

| Pfam   | Prosite   | Full Name                             | Description   |
|--------|-----------|---------------------------------------|---|
|        |           |                                       | <p>Baumelster W.<br/>J. Bacteriol. 176:1224-1233(1994).</p> <p>[3]<br/>Lemaire M., Ohayon H., Gounon P., Fujino T., Beguin P.<br/>J. Bacteriol. 177:2451-2459(1995).</p>  |
| Smr    |           | Smr domain                            | <p>Accession number: PF01713<br/>Definition: Smr domain<br/>Author: Bateman A<br/>Alignment method of seed: Clustalw<br/>Source of seed members: [1]<br/>Gathering cutoffs: 0 0<br/>Trusted cutoffs: 1.40 1.40<br/>Noise cutoffs: -7.90 -7.90<br/>HMM build command line: hmmbuild HMM SEED<br/>HMM build command line: hmmscalibrate --seed 0 HMM<br/>Reference Number: [1]<br/>Reference Medline: 10431172<br/>Reference Title: Smr: a bacterial and eukaryotic homologue of the C-terminal<br/>Reference Title: region of the MutS2 family.<br/>Reference Author: Moreira D, Philippe H;<br/>Reference Location: Trends Biochem Sci 1999;24:298-300.<br/>Database Reference: INTERPRO; IPR002625;<br/>Comment: This family includes the Smr (Small MutS Related) proteins,<br/>Comment: and the C-terminal region of the MutS2 protein. It has been<br/>Comment: suggested that this domain interacts with the MutS1<br/>Comment: Swiss:P23909 protein in the case of Smr proteins and with<br/>Comment: the N-terminal MutS related region of MutS2<br/>Swiss:P94545 [1].<br/>Number of members: 14</p>   |
| SRF-TF | PDOC00302 | MADS-box domain signature and profile | <p>A number of transcription factors contain a conserved domain of 56 amino-acid residues, sometimes known as the MADS-box domain [E1]. They are listed below:</p> <ul style="list-style-type: none"> <li>- Serum response factor (SRF) [1], a mammalian transcription factor that binds to the Serum Response Element (SRE). This is a short sequence of dyad symmetry located 300 bp to the 5' end of the transcription initiation site of genes such as c-fos.</li> <li>- Mammalian myocyte-specific enhancer factors 2A to 2D (MEF2A to MEF2D). These proteins are transcription factor which binds specifically to the MEF2 element present in the regulatory regions of many muscle-specific genes.</li> <li>- Drosophila myocyte-specific enhancer factor 2 (MEF2).</li> <li>- Yeast GRM/PRTF protein (gene MCM1) [2], a transcriptional regulator of mating-type-specific genes.</li> <li>- Yeast arginine metabolism regulation protein I (gene ARGR1 or ARG80).</li> <li>- Yeast transcription factor RLM1.</li> <li>- Yeast transcription factor SMP1.</li> <li>- Arabidopsis thaliana agamous protein (AG) [3], a probable transcription factor involved in regulating genes that determines stamen and carpel development in wild-type flowers. Mutations in the AG gene result in the replacement of the stamens by petals and the carpels by a new flower.</li> <li>- Arabidopsis thaliana homeotic proteins Apetala1 (AP1), Apetala3 (AP3) and</li> </ul> |

004001" E0267960

1002

| Pfam  | Prosite | Full Name     | Description   |
|-------|---------|---------------|---|
|       |         |               | <p>Pistillata (PI) which act locally to specify the identity of the floral meristem and to determine sepal and petal development [4].</p> <ul style="list-style-type: none"> <li>- Antirrhinum majus and tobacco homeotic protein deficiencies (DEFA) and globosa (GLO) [5]. Both proteins are transcription factors involved in the genetic control of flower development. Mutations in DEFA or GLO cause the transformation of petals into sepals and of stamens into carpels.</li> <li>- Arabidopsis thaliana putative transcription factors AGL1 to AGL6 [6].</li> <li>- Antirrhinum majus morphogenetic protein DEF H33 (squamosa).</li> </ul> <p>In SRF, the conserved domain has been shown [1] to be involved in DNA-binding and dimerization. We have derived a pattern that spans the complete length of the domain. The profile also spans the length of the MADS-box.</p> <p>Description of pattern(s) and/or profile(s)</p> <p>Consensus pattern R-x-[RK]-x(5)-I-x-[DNGSK]-x(3)-[KR]-x(2)-T-[FY]-x-[RK](3)-x(2)-[LIVM]-x-K(2)-A-x-E-[LIVM]-[STA]-x-L-x(4)-[LIVM]-x-[LIVM](3)-x(6)-[LIVMF]-x(2)-[FY]</p> <p>Sequences known to belong to this class detected by the pattern ALL.</p> <p>Other sequence(s) detected in SWISS-PROT NONE.</p> <p>Sequences known to belong to this class detected by the profile ALL.</p> <p>Other sequence(s) detected in SWISS-PROT NONE.</p> <p>Note this documentation entry is linked to both signature patterns and a profile. As the profile is much more sensitive than the patterns, you should use it if you have access to the necessary software tools to do so.</p> <p>Last update<br/>July 1999 / Pattern and text revised.</p> <p>References</p> <p>[ 1]<br/>Norman C., Runswick M., Pollock R., Treisman R.<br/>Cell 55:989-1003(1988).</p> <p>[ 2]<br/>Passmore S., Maine G.T., Eble R., Christ C., Tye B.-K.<br/>J. Mol. Biol. 204:593-606(1988).</p> <p>[ 3]<br/>Yanofsky M., Ma H., Bowman J., Drews G., Feldmann K.A., Meyerowitz E.M.<br/>Nature 346:35-39(1990).</p> <p>[ 4]<br/>Goto K., Meyerowitz E.M.<br/>Genes Dev. 8:1548-1560(1994).</p> <p>[ 5]<br/>Troebner W., Ramirez L., Motte P., Hue I., Huijser P., Loennig W.-E., Saedler H., Sommer H., Schwartz-Sommer Z.<br/>EMBO J. 11:4693-4704(1992).</p> <p>[ 6]<br/>Ma H., Yanofsky M.F., Meyerowitz E.M.<br/>Genes Dev. 5:484-495(1991).</p> <p>[E1]<br/><a href="http://transfac.gbf-braunschweig.de/cgi-bin/qt/getEntry.pl?C0014">http://transfac.gbf-braunschweig.de/cgi-bin/qt/getEntry.pl?C0014</a></p> |
| SRP19 |         | SRP19 protein | <p>Accession number: PF01922</p> <p>Definition: SRP19 protein</p> <p>Author: Enright A, Ouzounis C, Bateman A</p>   |

004007" E026" 960

1003

| Pfam | Prosite   | Full Name                                       | Description   |
|------|-----------|---|---|
|      |           |   | <p>Alignment method of seed: Clustalw<br/> Source of seed members: Enright A<br/> Gathering cutoffs: 25 25<br/> Trusted cutoffs: 31.20 31.20<br/> Noise cutoffs: -28.50 -28.50<br/> HMM build command line: hmmbuild -F HMM SEED<br/> HMM build command line: hmmscalibrate --seed 0 HMM<br/> Reference Number: [1]<br/> Reference Medline: 89041541<br/> Reference Title: Isolation and characterization of a cDNA clone encoding the<br/> Reference Title: 19 kDa protein of signal recognition particle (SRP):<br/> Reference Title: expression and binding to TSL RNA.<br/> Reference Author: Lingelbach K, Zwieb C, Webb JR, Marshallsay C, Hoben PJ,<br/> Reference Author: Walter P, Dobberstein B;<br/> Reference Location: Nucleic Acids Res 1988;16:9431-9442.<br/> Reference Number: [2]<br/> Reference Medline: 92220168<br/> Reference Title: SEC65 gene product is a subunit of the yeast signal<br/> Reference Title: recognition particle required for its integrity.<br/> Reference Author: Hann BC, Stirling CJ, Walter P;<br/> Reference Location: Nature 1992;356:532-533.<br/> Reference Number: [3]<br/> Reference Medline: 92220169<br/> Reference Title: The <i>S. cerevisiae</i> SEC65 gene encodes a component of yeast<br/> Reference Title: signal recognition particle with homology to human SRP19.<br/> Reference Author: Stirling CJ, Hewitt EW;<br/> Reference Location: Nature 1992;356:534-537.<br/> Database Reference INTERPRO; IPR002778;<br/> Comment: The signal recognition particle (SRP) binds to the signal peptide of<br/> Comment: proteins as they are being translated. The binding of the SRP halts<br/> Comment: translation and the complex is then transported to the endoplasmic<br/> Comment: reticulum's cytoplasmic surface. The SRP then aids translocation of<br/> Comment: the protein through the ER membrane. The SRP is a ribonucleoprotein<br/> Comment: that is composed of a small RNA and several proteins. One of these<br/> Comment: proteins is the SRP19 protein [1] (Sec65 in yeast [2,3]).<br/> Number of members: 13</p> |
| SSB  | PDOC00602 | Single-strand binding protein family signatures | <p>The <i>Escherichia coli</i> single-strand binding protein [1] (gene <i>ssb</i>), also known as the helix-destabilizing protein, is a protein of 177 amino acids. It binds tightly, as a homotetramer, to single-stranded DNA (ss-DNA) and plays an important role in DNA replication, recombination and repair.</p> <p>Closely related variants of SSB are encoded in the genome of a variety of large self-transmissible plasmids. SSB has also been characterized in bacteria such as <i>Proteus mirabilis</i> or <i>Serratia marcescens</i>.</p> <p>Eukaryotic mitochondrial proteins that bind ss-DNA and are probably involved in mitochondrial DNA replication are structurally and evolutionary related to prokaryotic SSB. Proteins currently known to belong to this subfamily are listed below [2].</p> <ul style="list-style-type: none"> <li>- Mammalian protein Mt-SSB (P16).</li> <li>- <i>Xenopus</i> Mt-SSBs and Mt-SSBr.</li> </ul>   |

004001 "E0264960

| Pfam         | Prosite | Full Name         | Description  |
|--------------|---------|-------------------|--|
|              |         |                   | <p>- Drosophila MtSSB.<br/>- Yeast protein RIM1.</p> <p>We have developed two signature patterns for these proteins. The first is a conserved region in the N-terminal section of the SSB's. The second is a centrally located region which, in Escherichia coli SSB, is known to be involved in the binding of DNA.</p> <p>Description of pattern(s) and/or profile(s)</p> <p>Consensus pattern [LIVMF]-[NST]-[KRHST]-[LIVM]-x-[LIVMF](2)-G-[NHRK]-[LIVMA]-[GST]-x-[DENT]<br/>Sequences known to belong to this class detected by the pattern ALL.<br/>Other sequence(s) detected in SWISS-PROT NONE.</p> <p>Consensus pattern T-x-W-[HY]-[RNS]-[LIVM]-x-[LIVMF]-[FY]-[NGKR]<br/>Sequences known to belong to this class detected by the pattern A majority.<br/>Other sequence(s) detected in SWISS-PROT NONE.<br/>Last update<br/>December 1999 / Patterns and text revised.<br/>References<br/>[ 1]<br/>Meyer R.R., Laine P.S.<br/>Microbiol. Rev. 54:342-380(1990).<br/>[ 2]<br/>Stroumbakis N.D., Li Z., Tolias P.P.<br/>Gene 143:171-177(1994).</p> |
| START        |         | START domain      | <p>Accession number: PF01852<br/>Definition: START domain<br/>Author: SMART<br/>Alignment method of seed: Manual<br/>Source of seed members: Alignment kindly provided by SMART<br/>Gathering cutoffs: 25 25<br/>Trusted cutoffs: 106.20 106.20<br/>Noise cutoffs: -20.90 -20.90<br/>HMM build command line: hmmbuild HMM SEED<br/>HMM build command line: hmmcalibrate --seed 0 HMM<br/>Reference Number: [1]<br/>Reference Medline: 99257451<br/>Reference Title: START: a lipid-binding domain in StAR, HD-ZIP and<br/>Reference Title: signalling proteins.<br/>Reference Author: Ponting CP, Aravind L;<br/>Reference Location: Trends Biochem Sci 1999;24:130-132.<br/>Database reference: SMART; START;<br/>Database Reference: INTERPRO; IPR002913;<br/>Number of members: 41</p>  |
| Sterol_desat |         | Sterol desaturase | <p>Accession number: PF01598<br/>Definition: Sterol desaturase<br/>Author: Bateman A<br/>Alignment method of seed: Clustalw<br/>Source of seed members: Pfam-B_905 (release 4.1)<br/>Gathering cutoffs: -13 -13<br/>Trusted cutoffs: 12.90 12.90<br/>Noise cutoffs: -44.50 -44.50<br/>HMM build command line: hmmbuild -F HMM SEED<br/>HMM build command line: hmmcalibrate --seed 0 HMM<br/>Reference Number: [1]<br/>Reference Medline: 91323727<br/>Reference Title: Cloning, disruption and sequence of the gene encoding yeast<br/>Reference Title: C-5 sterol desaturase.<br/>Reference Author: Arthington BA, Bennett LG, Skatrud PL,</p>   |

| Pfam      | Prosite   | Full Name             | Description   |
|-----------|-----------|-----------------------|---|
|           |           |                       | <p>Guynn CJ, Barbuch<br/> Reference Author: RJ, Ulbright CE, Bard M;<br/> Reference Location: Gene 1991;102:39-44.<br/> Reference Number: [2]<br/> Reference Medline: 96133902<br/> Reference Title: Cloning and characterization of ERG25, the <i>Saccharomyces cerevisiae</i> gene encoding C-4 sterol methyl oxidase.<br/> Reference Author: Bard M, Bruner DA, Pierson CA, Lees ND, Biermann B, Frye L,<br/> Reference Author: Koegel C, Barbuch R;<br/> Reference Location: Proc Natl Acad Sci U S A 1996;93:186-190.<br/> Reference Number: [3]<br/> Reference Medline: 96351930<br/> Reference Title: Molecular characterization of the CER1 gene of <i>arabidopsis</i><br/> Reference Title: involved in epicuticular wax biosynthesis and pollen fertility.<br/> Reference Author: Aarts MG, Keijzer CJ, Stiekema WJ, Pereira A;<br/> Reference Location: Plant Cell 1995;7:2115-2127.<br/> Database Reference: INTERPRO; IPR001541;<br/> Database reference: PFAMB; PB041851;<br/> Comment: This family includes C-5 sterol desaturase and C-4 sterol methyl oxidase. Members of this family are involved in cholesterol biosynthesis and biosynthesis a plant cuticular wax.<br/> Comment: These enzymes contain many conserved histidine residues. Members of this family are integral membrane proteins.<br/> Comment: Number of members: 34</p>   |
| Sulfatase | PDOC00117 | Sulfatases signatures | <p>Sulfatases (EC 3.1.6.-) are enzymes that hydrolyze various sulfate esters. The sequence of different types of sulfatases are available. These enzymes are:</p> <ul style="list-style-type: none"> <li>- Arylsulfatase A (EC 3.1.6.8) (ASA), a lysosomal enzyme which hydrolyzes cerebroside sulfate.</li> <li>- Arylsulfatase B (EC 3.1.6.12) (ASB), a lysosomal enzyme which hydrolyzes the sulfate ester group from N-acetylgalactosamine 4-sulfate residues of dermatan sulfate.</li> <li>- Arylsulfatase C (ASD).</li> <li>- Arylsulfatase E (ASE).</li> <li>- Steryl-sulfatase (EC 3.1.6.2) (STS) (arylsulfatase C), a membrane bound microsomal enzyme which hydrolyzes 3-beta-hydroxy steroid sulfates.</li> <li>- Iduronate 2-sulfatase precursor (EC 3.1.6.13) (IDS), a lysosomal enzyme that hydrolyzes the 2-sulfate groups from non-reducing-terminal iduronic acid residues in dermatan sulfate and heparan sulfate.</li> <li>- N-acetylgalactosamine-6-sulfatase (EC 3.1.6.4), an enzyme that hydrolyzes the 6-sulfate groups of the N-acetyl-D-galactosamine 6-sulfate units of chondroitin sulfate and the D-galactose 6-sulfate units of keratan sulfate.</li> <li>- Choline sulfatase (EC 3.1.6.6) (gene betC), a bacterial enzyme that converts choline-O-sulfate to choline.</li> <li>- Glucosamine-6-sulfatase (EC 3.1.6.14) (G6S), a lysosomal enzyme that hydrolyzes the N-acetyl-D-glucosamine 6-sulfate units of heparan sulfate</li> </ul> |

004001" E0262960

| Pfam           | Prosite   | Full Name                      | Description   |
|----------------|-----------|--------------------------------|---|
|                |           |                                | <p>and keratan sulfate.</p> <ul style="list-style-type: none"> <li>- N-sulphoglucosamine sulphonydrolase (EC 3.10.1.1) (sulphamidase), the lysosomal enzyme that catalyzes the hydrolysis of N-sulfo-d-glucosamine into glucosamine and sulfate.</li> <li>- Sea urchin embryo arylsulfatase (EC 3.1.6.1).</li> <li>- Green alga arylsulfatase (EC 3.1.6.1), an enzyme which plays an important role in the mineralization of sulfates.</li> <li>- Arylsulfatase (EC 3.1.6.1) from <i>Escherichia coli</i> (gene <i>asIA</i>), <i>Klebsiella aerogenes</i> (gene <i>atsA</i>) and <i>Pseudomonas aeruginosa</i> (gene <i>atsA</i>).</li> <li>- <i>Escherichia coli</i> hypothetical protein <i>ydJ</i>.</li> </ul> <p>It has been shown that all these sulfatases are structurally related [1,2,3].</p> <p>As signature patterns for that family of enzymes we have selected the two best conserved regions. Both regions are located in the N-terminal section of these enzymes. The first region contains a conserved arginine which could be implicated in the catalytic mechanism; it is located four residues after a position that, in eukaryotic sulfatases, is a conserved cysteine which has been shown [4] to be modified to 2-amino-3-oxopropionic acid. In prokaryotes, this cysteine is replaced by a serine.</p> <p>Description of pattern(s) and/or profile(s)</p> <p>Consensus pattern [SAP]-[LIVMST]-[CS]-[STAC]-P-[STA]-R-x(2)-[LIVFW](2)-[TAR]-G [R is a putative active site residue]<br/>Sequences known to belong to this class detected by the pattern ALL.<br/>Other sequence(s) detected in SWISS-PROT NONE.</p> <p>Consensus pattern G-[YV]-x-[ST]-x(2)-[IVAS]-G-K-x(0,1)-[FYWMK]-[HL]<br/>Sequences known to belong to this class detected by the pattern ALL.<br/>Other sequence(s) detected in SWISS-PROT NONE.</p> <p>Last update<br/>December 1999 / Patterns and text revised.</p> <p>References<br/>[ 1 ]<br/>Peters C., Schmidt B., Rommerskirch W., Rupp K., Zuehlisdorf M., Vingron M., Meyer H.E., Pohlmann R., von Figura K.<br/><i>J. Biol. Chem.</i> 265:3374-3381(1990).</p> <p>[ 2 ]<br/>Wilson P.J., Morris C.P., Anson D.S., Occhiodoro T., Bielicki J., Clements P.R., Hopwood J.J.<br/><i>Proc. Natl. Acad. Sci. U.S.A.</i> 87:8531-8535(1990).</p> <p>[ 3 ]<br/>de Hostos E.L., Schilling J., Grossman A.R.<br/><i>Mol. Gen. Genet.</i> 218:229-239(1989).</p> <p>[ 4 ]<br/>Selmer T., Hallmann A., Schmidt B., Sumper M., von Figura K.<br/><i>Eur. J. Biochem.</i> 238:341-345(1996).</p> |
| Sulfate_transp | PDOC00870 | Sulfate transporters signature | <p>A number of proteins involved in the transport of sulfate across a membrane as well as some yet uncharacterized proteins have been shown [1,2] to be evolutionary related. These proteins are:</p>   |

004001 00262960

| Pfam      | Prosite | Full Name | Description   |
|-----------|---------|-----------|---|
|           |         |           | <ul style="list-style-type: none"> <li>- Neurospora crassa sulfate permease II (gene cys-14).</li> <li>- Yeast sulfate permeases (genes SUL1 and SUL2).</li> <li>- Rat sulfate anion transporter 1 (SAT-1).</li> <li>- Mammalian DTDST, a probable sulfate transporter which, in Human, is involved in the genetic disease, diastrophic dysplasia (DTD).</li> <li>- Sulfate transporters 1, 2 and 3 from the legume Stylosanthes hamata.</li> <li>- Human pendrin (gene PDS), which is involved in a number of hearing loss genetic diseases.</li> <li>- Human protein DRA (Down-Regulated in Adenoma).</li> <li>- Soybean early nodulin 70.</li> <li>- Escherichia coli hypothetical protein ychM.</li> <li>- Caenorhabditis elegans hypothetical protein F41D9.5.</li> </ul> <p>As expected by their transport function, these proteins are highly hydrophobic and seem to contain about 12 transmembrane domains. The best conserved region seems to be located in the second transmembrane region and is used as a signature pattern.</p> <p>Description of pattern(s) and/or profile(s)</p> <p>Consensus pattern [PAV]-x-Y-[GS]-L-Y-[STAG](2)-x(4)-[LIVFYA]-[LIVST]-[YI]-x(3)-[GA]-[GST]-S-[KR]<br/> Sequences known to belong to this class detected by the pattern ALL.<br/> Other sequence(s) detected in SWISS-PROT NONE.<br/> Last update July 1999 / Pattern and text revised.<br/> References<br/> [1]<br/> Sandal N.N., Marcker K.A.<br/> Trends Biochem. Sci. 19:19-19(1994).<br/> [2]<br/> Smith F.W., Hawkesford M.J., Prosser I.M., Clarkson D.T.<br/> Mol. Gen. Genet. 247:709-715(1995).</p> |
| Synuclein |         | Synuclein | <p>Accession number: PF01387<br/> Definition: Synuclein<br/> Author: Bateman A<br/> Alignment method of seed: Clustalw<br/> Source of seed members: [1]<br/> Gathering cutoffs: 25 25<br/> Trusted cutoffs: 197.80 197.80<br/> Noise cutoffs: -33.80 -33.80<br/> HMM build command line: hmmbuild HMM SEED<br/> HMM build command line: hmmscalibrate --seed 0 HMM<br/> Reference Number: [1]<br/> Reference Medline: 98424410<br/> Reference Title: The synuclein family.<br/> Reference Author: Lavedan C;<br/> Reference Location: Genome Res 1998;8:871-880.<br/> Database Reference INTERPRO; IPR001058;<br/> Comment: There are three types of synucleins in humans, these<br/> Comment: are called alpha, beta and gamma. Alpha<br/> Comment: synuclein has<br/> Comment: been found mutated in families with<br/> Comment: autosomal dominant<br/> Comment: Parkinson's disease. A peptide of alpha<br/> Comment: synuclein has<br/> Comment: also been found in amyloid plaques in<br/> Comment: Alzheimer's<br/> Comment: patients.<br/> Number of members: 12</p>  |

004001" E0264960

| Pfam | Prosite   | Full Name                       | Description   |
|------|-----------|---------------------------------|---|
| TEA  | PDOC00479 | TEA domain signature            | <p>The TEA domain [1,E1] is a DNA-binding region of about 66 to 68 amino acids which has been found in the N-terminal section of the following nuclear regulatory proteins:</p> <ul style="list-style-type: none"> <li>- Mammalian enhancer factor TEF-1. TEF-1 can bind to two distinct sequences in the SV40 enhancer and is a transcriptional activator.</li> <li>- Mammalian TEF-3, TEF-4 and TEF-5 [2], putative transcriptional activators highly similar to TEF-1.</li> <li>- Drosophila scalloped protein (gene sd), a probable transcription factor that functions in the regulation of cell-specific gene expression during Drosophila development, particularly in the differentiation of the nervous system [3].</li> <li>- Emericella nidulans regulatory protein abaA. AbaA is involved in the regulation of conidiation (asexual spore); its expression leads to the cessation of vegetative growth.</li> <li>- Yeast trans-acting factor TEC1. TEC1 is involved in the activation of the Ty1 retrotransposon.</li> <li>- Caenorhabditis elegans hypothetical protein F28B12.2.</li> </ul> <p>As a signature pattern, we have used positions 39 to 67 of the TEA domain.</p> <p>Description of pattern(s) and/or profile(s)</p> <p>Consensus pattern G-R-N-E-L-I-x(2)-Y-I-x(3)-[TC]-x(3)-R-T-[RK](2)-Q-[LIVM]-S-S-H-[LIVM]-Q-V</p> <p>Sequences known to belong to this class detected by the pattern ALL.</p> <p>Other sequence(s) detected in SWISS-PROT NONE.</p> <p>Last update November 1997 / Pattern and text revised.</p> <p>References</p> <p>[ 1 ]<br/>Buerklin T.R.<br/>Cell 66:11-12(1991).</p> <p>[ 2 ]<br/>Jacquemin P., Hwang J.-J., Martial J.A., Dolle P., Davidson I.<br/>J. Biol. Chem. 271:21775-21785(1996).</p> <p>[ 3 ]<br/>Campbell S.D., Inamdar M., Rodrigues V., Raghavan V., Palazzolo M., Chovnick A.<br/>Genes Dev. 6:367-379(1992).</p> <p>[E1]<br/><a href="http://transfac.gbf-braunschweig.de/cgi-bin/qt/getEntry.pl?C0024">http://transfac.gbf-braunschweig.de/cgi-bin/qt/getEntry.pl?C0024</a></p> |
| TGT  |           | Queuine tRNA-ribosyltransferase | <p>Accession number: PF01702</p> <p>Definition: Queuine tRNA-ribosyltransferase</p> <p>Author: Bashon M, Bateman A</p> <p>Alignment method of seed: Clustalw</p> <p>Source of seed members: Pfam-B_1643 (release 4.1)</p> <p>Gathering cutoffs: -132 -132</p> <p>Trusted cutoffs: -110.00 -110.00</p> <p>Noise cutoffs: -155.40 -155.40</p> <p>HMM build command line: hmmbuild -F HMM SEED</p> <p>HMM build command line: hmmscalibrate --seed 0 HMM</p> <p>Reference Number: [1]</p> <p>Reference Medline: 96256303</p> <p>Reference Title: Crystal structure of tRNA-guanine transglycosylase: RNA</p>   |

004001" E0262950



| Pfam | Prosite | Full Name   | Description   |
|------|---------|-------------|---|
|      |         |             | <p>Reference Title: modification by base exchange.<br/> Reference Author: Romier C, Reuter K, Suck D, Ficner R;<br/> Reference Location: EMBO J 1996;15:2850-2857.<br/> Reference Number: [2]<br/> Reference Medline: 93287116<br/> Reference Title: tRNA-guanine transglycosylase from<br/> Escherichia coli.<br/> Reference Title: Overexpression, purification and quaternary<br/> structure.<br/> Reference Author: Garcia GA, Koch KA, Chong S;<br/> Reference Location: J Mol Biol 1993;231:489-497.<br/> Database Reference: SCOP; 1pud; fa; [SCOP-USA][CATH-<br/> PDBSUM]<br/> Database Reference INTERPRO; IPR002616;<br/> Database Reference PDB; 1efz A; 138; 379;<br/> Database Reference PDB; 1enu A; 138; 379;<br/> Database Reference PDB; 1pud ; 138; 379;<br/> Database Reference PDB; 1wkd ; 138; 379;<br/> Database Reference PDB; 1wke ; 138; 379;<br/> Database Reference PDB; 1wkf ; 138; 379;<br/> Database reference: PFAMB; PB037884;<br/> Comment: This is a family of queuine tRNA-<br/> ribosyltransferases<br/> Comment: EC:2.4.2.29, also known as tRNA-guanine<br/> transglycosylase<br/> Comment: and guanine insertion enzyme.<br/> Comment: Queuine tRNA-ribosyltransferase modifies<br/> tRNAs for asparagine,<br/> Comment: aspartic acid, histidine and tyrosine with<br/> queuine.<br/> Comment: It catalyses the exchange of guanine-34 at<br/> the wobble position with<br/> Comment: 7-aminomethyl-7-deazaguanine, and the<br/> addition of a cyclopentenediol<br/> Comment: moiety to 7-aminomethyl-7-deazaguanine-<br/> 34 tRNA; giving a hypermodified<br/> Comment: base queuine in the wobble position [1,2].<br/> Comment: The aligned region contains a zinc binding<br/> motif C-x-C-x2-C-x29-H,<br/> Comment: and important tRNA and 7-aminomethyl-<br/> 7deazaguanine binding residues [1].<br/> Number of members: 24</p> |
| Thi4 |         | Thi4 family | <p>Accession number: PF01946<br/> Definition: Thi4 family<br/> Author: Enright A, Ouzounis C, Bateman A<br/> Alignment method of seed: Clustalw<br/> Source of seed members: Enright A<br/> Gathering cutoffs: 25 25<br/> Trusted cutoffs: 526.80 526.80<br/> Noise cutoffs: -105.00 -105.00<br/> HMM build command line: hmmbuild -F HMM SEED<br/> HMM build command line: hmmscalibrate --seed 0 HMM<br/> Reference Number: [1]<br/> Reference Medline: 95050223<br/> Reference Title: Cloning, nucleotide sequence, and<br/> regulation of<br/> Reference Title: Schizosaccharomyces pombe thi4, a<br/> thiamine biosynthetic<br/> Reference Title: gene.<br/> Reference Author: Zurlinden A, Schweingruber ME;<br/> Reference Location: J Bacteriol 1994;176:6631-6635.<br/> Database Reference INTERPRO; IPR002922;<br/> Comment: This family includes Swiss:P32318 a<br/> putative thiamine biosynthetic<br/> Comment: enzyme.<br/> Number of members: 14</p>   |
| ThiC |         | ThiC family | <p>Accession number: PF01964<br/> Definition: ThiC family<br/> Author: Enright A, Ouzounis C, Bateman A<br/> Alignment method of seed: Clustalw<br/> Source of seed members: Enright A<br/> Gathering cutoffs: 25 25</p>  |

004007" 00264960

| Pfam | Prosite | Full Name        | Description   |
|------|---------|------------------|---|
|      |         |                  | <p>Trusted cutoffs: 1047.20 1047.20<br/> Noise cutoffs: -338.20 -338.20<br/> HMM build command line: hmmbuild -F HMM SEED<br/> HMM build command line: hmmscalibrate --seed 0 HMM<br/> Reference Number: [1]<br/> Reference Medline: 93163063<br/> Reference Title: Structural genes for thiamine biosynthetic enzymes<br/> Reference Title: (thiCEFGH) in Escherichia coli K-12.<br/> Reference Author: Vander Horn PB, Backstrom AD, Stewart V, Begley TP;<br/> Reference Location: J Bacteriol 1993;175:982-992.<br/> Reference Number: [2]<br/> Reference Medline: 99311269<br/> Reference Title: Thiamin biosynthesis in prokaryotes.<br/> Reference Author: Begley TP, Downs DM, Ealick SE, McLafferty FW, Van Loon AP,<br/> Reference Author: Taylor S, Campobasso N, Chiu HJ, Kinsland C, Reddick JJ, Xi<br/> Reference Author: J;<br/> Reference Location: Arch Microbiol 1999;171:293-300.<br/> Reference Number: [3]<br/> Reference Medline: 97284509<br/> Reference Title: Characterization of the Bacillus subtilis thiC operon<br/> Reference Title: involved in thiamine biosynthesis.<br/> Reference Author: Zhang Y, Taylor SV, Chiu HJ, Begley TP;<br/> Reference Location: J Bacteriol 1997;179:3030-3035.<br/> Database Reference INTERPRO: IPR002817;<br/> Comment: ThiC is found within the thiamine biosynthesis operon. ThiC is<br/> Comment: involved in pyrimidine biosynthesis [2].<br/> Comment: ThiC catalyzes the substitution of the pyrophosphate of<br/> Comment: 2-methyl-4-amino-5-hydroxymethylpyrimidine pyrophosphate by<br/> Comment: 4-methyl-5-(beta-hydroxyethyl)thiazole phosphate to yield thiamine<br/> Comment: phosphate [3].<br/> Number of members: 12</p> |
| ThiJ |         | ThiJ/Pfpl family | <p>Accession number: PF01965<br/> Definition: ThiJ/Pfpl family<br/> Author: Enright A, Ouzounis C, Bateman A<br/> Alignment method of seed: Clustalw<br/> Source of seed members: Enright A<br/> Gathering cutoffs: -40.2 -40.2<br/> Trusted cutoffs: -40.20 -40.20<br/> Noise cutoffs: -47.00 -47.00<br/> HMM build command line: hmmbuild HMM SEED<br/> HMM build command line: hmmscalibrate --seed 0 HMM<br/> Reference Number: [1]<br/> Reference Medline: 97039868<br/> Reference Title: The thiJ locus and its relation to phosphorylation of<br/> Reference Title: hydroxymethylpyrimidine in Escherichia coli.<br/> Reference Author: Mizote T, Tsuda M, Nakazawa T, Nakayama H;<br/> Reference Location: Microbiology 1996;142:2969-2974.<br/> Reference Number: [2]<br/> Reference Medline: 96196168<br/> Reference Title: Sequence, expression in Escherichia coli, and analysis of<br/> Reference Title: the gene encoding a novel intracellular protease (Pfpl)<br/> Reference Title: from the hyperthermophilic archaeon Pyrococcus furiosus.<br/> Reference Author: Halio SB, Blumentals II, Short SA, Merrill BM, Kelly RM;<br/> Reference Location: J Bacteriol 1996;178:2605-2612.<br/> Database Reference INTERPRO: IPR002818;<br/> Database reference: PFAMB; PB002774;<br/> Database reference: PFAMB; PB007213;</p>  |

004007 "E026X960

| Pfam            | Prosite   | Full Name                                  | Description   |
|-----------------|-----------|--|---|
|                 |           |  | <p>Database reference: PFAMB; PB041784;<br/> Comment: This family includes ThiJ a thiamine biosynthesis<br/> Comment: enzyme [1] that catalyses the phosphorylation of<br/> Comment: hydroxymethylpyrimidine (HMP) to HMP monophosphate EC:2.7.1.49.<br/> Comment: The family also includes a the protease PfpI<br/> Swiss:Q51732 [2].<br/> Number of members: 34</p>   |
| Thr_dehydrat_C  |           | C-terminal domain of Threonine dehydratase | <p>Accession number: PF00585<br/> Definition: C-terminal domain of Threonine dehydratase<br/> Previous Pfam IDs: Thr_dehydratase_C;<br/> Author: Bateman A<br/> Alignment method of seed: Clustalw<br/> Source of seed members: Bateman A<br/> Gathering cutoffs: 25 25<br/> Trusted cutoffs: 99.90 51.30<br/> Noise cutoffs: -1.10 -1.10<br/> HMM build command line: hmmbuild HMM SEED<br/> HMM build command line: hmmscalibrate --seed 0 HMM<br/> Reference Number: [1]<br/> Reference Medline: 98230745<br/> Reference Title: Structure and control of pyridoxal phosphate dependent<br/> Reference Title: allosteric threonine deaminase.<br/> Reference Author: Gallagher DT, Gilliland GL, Xiao G, Zondlo J, Fisher KE,<br/> Reference Author: Chinchilla D, Eisenstein E;<br/> Reference Location: Structure 1998;6:465-475.<br/> Database Reference: SCOP; 1tdj; fa; [SCOP-USA][CATH-PDBSUM]<br/> Database Reference INTERPRO; IPR001721;<br/> Database Reference PDB; 1tdj ; 424; 512;<br/> Database Reference PDB; 1tdj ; 329; 419;<br/> Comment: -I- Threonine dehydratases PALP all contain a carboxy<br/> Comment: terminal region. This region may have a regulatory role.<br/> Comment: Some members contain two copies of this region.<br/> Number of members: 30</p> |
| thymidylat_synt | PDOC00086 | Thymidylate synthase active site           | <p>Thymidylate synthase (EC 2.1.1.45) [1,2] catalyzes the reductive methylation of dUMP to dTMP with concomitant conversion of 5,10-methylenetetrahydrofolate to dihydrofolate. Thymidylate synthase plays an essential role in DNA synthesis and is an important target for certain chemotherapeutic drugs.</p> <p>Thymidylate synthase is an enzyme of about 30 to 35 Kd in most species except in protozoan and plants where it exists as a bifunctional enzyme that includes a dihydrofolate reductase domain.</p> <p>A cysteine residue is involved in the catalytic mechanism (it covalently binds the 5,6-dihydro-dUMP intermediate). The sequence around the active site of this enzyme is conserved from phages to vertebrates.</p> <p>Description of pattern(s) and/or profile(s)</p> <p>Consensus pattern R-x(2)-[LIVM]-x(3)-[FW]-[QN]-x(8,9)-[LV]-x-P-C-[HAVM]-x(3)-[QMT]-[FYW]-x-[LV] [C is the active site residue]<br/> Sequences known to belong to this class detected by the pattern ALL.<br/> Other sequence(s) detected in SWISS-PROT NONE.</p>   |

004007" E0264960

| Pfam         | Prosite   | Full Name                                   | Description   |
|--------------|-----------|---|---|
|              |           |   | <p>Last update<br/>November 1997 / Pattern and text revised.</p> <p>References</p> <p>[ 1 ]<br/>Benkovic S.J.<br/>Annu. Rev. Biochem. 49:227-251(1980).</p> <p>[ 2 ]<br/>Ross P., O'Gara F., Condon S.<br/>Appl. Environ. Microbiol. 56:2156-2163(1990).</p>  |
| Top6A        |           | Type II DNA topoisomerase                   | <p>Accession number: PF01962</p> <p>Definition: Type II DNA topoisomerase</p> <p>Author: Enright A, Ouzounis C, Bateman A</p> <p>Alignment method of seed: Clustalw</p> <p>Source of seed members: Enright A</p> <p>Gathering cutoffs: -99 -99</p> <p>Trusted cutoffs: -40.40 -40.40</p> <p>Noise cutoffs: -158.40 -158.40</p> <p>HMM build command line: hmmbuild -F HMM SEED</p> <p>HMM build command line: hmmcalibrate --seed 0 HMM</p> <p>Reference Number: [1]</p> <p>Reference Medline: 97238688</p> <p>Reference Title: An atypical topoisomerase II from Archaea with implications for meiotic recombination [see comments]</p> <p>Reference Author: Bergerat A, de Massy B, Gadelle D, Varoutas PC, Nicolas A,</p> <p>Reference Author: Forterre P;</p> <p>Reference Location: Nature 1997;386:414-417.</p> <p>Database Reference: SCOP; 1d3y; fa; [SCOP-USA][CATH-PDBSUM]</p> <p>Database Reference INTERPRO; IPR002815;</p> <p>Database Reference PDB; 1d3y A; 77; 363;</p> <p>Database Reference PDB; 1d3y B; 77; 363;</p> <p>Comment: Members of this family are the A subunit from type II DNA</p> <p>Comment: topoisomerases. Type II DNA</p> <p>Comment: topoisomerases catalyse the relaxation</p> <p>Comment: of DNA supercoiling by causing transient double strand breaks.</p> <p>Comment: The family includes topoisomerase VI subunit A from archaeobacteria</p> <p>Comment: Swiss:Q57815 EC:5.99.1.3 and SPO11 from yeast Swiss:P23179.</p> <p>Comment: A conserved tyrosine is thought to be involved in breaking the</p> <p>Comment: double stranded DNA [1].</p> <p>Number of members: 9</p> |
| Topoisom_bac | PDOC00333 | Prokaryotic DNA topoisomerase I active site | <p>DNA topoisomerase I (EC 5.99.1.2) [1,2,3,4,E1] is one of the two types of enzyme that catalyze the interconversion of topological DNA isomers. Type I topoisomerases act by catalyzing the transient breakage of DNA, one strand at a time, and the subsequent rejoining of the strands. When a prokaryotic type I topoisomerase breaks a DNA backbone bond, it simultaneously forms a protein-DNA link where the hydroxyl group of a tyrosine residue is joined to a 5'-phosphate on DNA, at one end of the enzyme-severed DNA strand.</p> <p>Prokaryotic organisms, such as Escherichia coli, have two type I topoisomerase isozymes: topoisomerase I (gene topA) and topoisomerase III (gene topB). Eukaryotes also contain homologs of prokaryotic topoisomerase III.</p> <p>There are a number of conserved residues in the region around the active site</p>   |

004007 00264960

| Pfam     | Prosite | Full Name                  | Description  |
|----------|---------|----------------------------|--|
|          |         |                            | <p>tyrosine; we used this region as a signature pattern.</p> <p>Description of pattern(s) and/or profile(s)</p> <p>Consensus pattern [EQ]-x-L-Y-[DEQST]-x(3,12)-[LIV]-[ST]-Y-x-R-[ST]-[DEQS] [The second Y is the active site tyrosine]<br/>Sequences known to belong to this class detected by the pattern ALL.<br/>Other sequence(s) detected in SWISS-PROT NONE.<br/>Last update<br/>December 1999 / Pattern and text revised.<br/>References<br/>[ 1]<br/>Sternglanz R.<br/>Curr. Opin. Cell Biol. 1:533-535(1990).<br/>[ 2]<br/>Sharma A., Mondragon A.<br/>Curr. Opin. Struct. Biol. 5:39-47(1995).<br/>[ 3]<br/>Bjornsti M.-A.<br/>Curr. Opin. Struct. Biol. 1:99-103(1991).<br/>[ 4]<br/>Roca J.<br/>Trends Biochem. Sci. 20:156-160(1995).<br/>[E1]<br/><a href="http://ellington.pharm.arizona.edu/~bear/top/topo.html">http://ellington.pharm.arizona.edu/~bear/top/topo.html</a></p>                 |
| toxin_3  |         | long chain scorpion toxins | <p>Accession number: PF00537<br/>Definition: long chain scorpion toxins<br/>Author: Bateman A<br/>Alignment method of seed: Manual<br/>Source of seed members: Arne Elofsson.<br/>Gathering cutoffs: 25 25<br/>Trusted cutoffs: 59.50 59.50<br/>Noise cutoffs: -3.80 -3.80<br/>HMM build command line: hmmbuild HMM SEED<br/>HMM build command line: hmmcalibrate --seed 0 HMM<br/>Database Reference: SCOP; 2sn3; fa; [SCOP-USA][CATH-PDBSUM]<br/>Database Reference INTERPRO; IPR002061;<br/>Comment: -!- Scorpion toxins bind to sodium channels and inhibit the activation<br/>Comment: mechanisms of the channels, thereby blocking neuronal transmission.<br/>Number of members: 77</p>  |
| Translin |         | Translin family            | <p>Accession number: PF01997<br/>Definition: Translin family<br/>Previous Pfam IDs: DUF130;<br/>Author: Enright A, Ouzounis C, Bateman A<br/>Alignment method of seed: Clustalw<br/>Source of seed members: Enright A<br/>Gathering cutoffs: 25 25<br/>Trusted cutoffs: 299.50 299.50<br/>Noise cutoffs: -72.40 -72.40<br/>HMM build command line: hmmbuild HMM SEED<br/>HMM build command line: hmmcalibrate --seed 0 HMM<br/>Reference Number: [1]<br/>Reference Medline: 97165975<br/>Reference Title: Isolation and characterization of a cDNA encoding a<br/>Reference Title: Translin-like protein, TRAX.<br/>Reference Author: Aoki K, Ishida R, Kasai M;<br/>Reference Location: FEBS Lett 1997;401:109-112.<br/>Database Reference INTERPRO; IPR002848;<br/>Comment: Members of this family include Translin<br/>Swiss:Q15631 that interacts<br/>Comment: with DNA and forms a ring around the DNA.</p> |

004001 00264960

| Pfam             | Prosite   | Full Name                            | Description   |
|------------------|-----------|--------------------------------------|---|
|                  |           |                                      | This family also includes<br>Comment: Swiss:Q99598, that was found to interact with translin with yeast<br>Comment: two-hybrid screen [1].<br>Number of members: 10   |
| Transposase_19   |           | Transposase 19                       | Members of this family are capable of in vitro and/or in vivo insertion of a donor polynucleotide into a target polynucleotide. Such biological activity is useful for inserting DNA into host genome, for example, for cloning purposes to generate a desired vector in vitro.   |
| TRANSPOSASE_IS30 | PDOC00801 | Transposases, IS30 family, signature | <p>Autonomous mobile genetic elements such as transposon or insertion sequences (IS) encode an enzyme, called transposase, required for excising and inserting the mobile element. On the basis of sequence similarities, transposases can be grouped into various families. One of these families has been shown [1,2] to consist of transposases from the following elements:</p> <ul style="list-style-type: none"> <li>- Is30 from Escherichia coli.</li> <li>- Is1086 from Alcaligenes eutrophus.</li> <li>- Is1161 from Streptococcus salivarius.</li> <li>- Is4351 (Tn4551) from Bacteroides fragilis.</li> </ul> <p>These transposases are proteins of 340 to 380 amino acids. The best conserved region is located in their C-terminal section and is used as a signature pattern.</p> <p>Description of pattern(s) and/or profile(s)</p> <p>Consensus pattern R-G-x(2)-E-N-x-N-G-[LIVM](2)-R-[QE]-[LIVMFY](2)-P-K<br/>Sequences known to belong to this class detected by the pattern ALL.<br/>Other sequence(s) detected in SWISS-PROT NONE.<br/>Last update<br/>November 1995 / First entry.<br/>References<br/>[ 1]<br/>Dong Q., Sadouk A., van der Lelie D., Taghavi S., Ferhat A., Nuyten J.M., Borremans B., Mergeay M., Toussaint A. J. Bacteriol. 174:8133-8138(1992).<br/>[ 2]<br/>Giffard P.M., Rathsam C., Kwan E., Kwan D.W.L., Bunny K.L., Koo S.-P., Jacques N.A. J. Gen. Microbiol. 139:913-920(1993).</p> |
| Transthyretin    | PDOC00617 | Transthyretin signatures             | <p>Transthyretin (prealbumin) [1] is a thyroid hormone-binding protein that seems to transport thyroxine (T4) from the bloodstream to the brain. It is a protein of about 130 amino acids that assembles as a homotetramer and forms an internal channel that binds thyroxine. Transthyretin is mainly synthesized in the brain choroid plexus. In humans, variants of the protein are associated with distinct forms of amyloidosis.</p> <p>The sequence of transthyretin is highly conserved in vertebrates. A number of uncharacterized proteins also belong to this family:</p> <ul style="list-style-type: none"> <li>- Escherichia coli hypothetical protein yedX.</li> <li>- Bacillus subtilis hypothetical protein yunM.</li> </ul>   |

004007 "E0264960

| Pfam | Prosite | Full Name                                      | Description  |
|------|---------|--|--|
|      |         |  | <p>- Caenorhabditis elegans hypothetical protein R09H10.3.<br/>- Caenorhabditis elegans hypothetical protein ZK697.8.</p> <p>We selected two regions as signature patterns. The first located in the N-terminal extremity starts with a lysine known to be involved in binding T4.<br/>The second pattern is located in the C-terminal extremity.</p> <p>Description of pattern(s) and/or profile(s)</p> <p>Consensus pattern [KH]-[IV]-L-[DN]-x(3)-G-x-P-A-x(2)-[IV]-x-[IV]<br/>[The K binds thyroxine]<br/>Sequences known to belong to this class detected by the pattern ALL.<br/>Other sequence(s) detected in SWISS-PROT NONE.</p> <p>Consensus pattern Y-[TH]-[IV]-[AP]-x(2)-L-S-[PQ]-[FYW]-[GS]-[FY]-[QS]<br/>Sequences known to belong to this class detected by the pattern ALL.<br/>Other sequence(s) detected in SWISS-PROT NONE.<br/>Last update<br/>July 1999 / Patterns and text revised.<br/>References<br/>[1]<br/>Schreiber G., Richardson S.J.<br/>Comp. Biochem. Physiol. 116B:137-160(1997).</p>  |
| TRM  |         | N2,N2-dimethylguanosine tRNA methyltransferase | <p>Accession number: PF02005<br/>Definition: N2,N2-dimethylguanosine tRNA methyltransferase<br/>Author: Enright A, Ouzounis C, Bateman A<br/>Alignment method of seed: Clustalw<br/>Source of seed members: Enright A<br/>Gathering cutoffs: 25 25<br/>Trusted cutoffs: 664.60 664.60<br/>Noise cutoffs: -259.50 -259.50<br/>HMM build command line: hmmbuild -F HMM SEED<br/>HMM build command line: hmmscalibrate --seed 0 HMM<br/>Reference Number: [1]<br/>Reference Medline: 98352211<br/>Reference Title: The tRNA(guanine-26,N2-N2) methyltransferase (Trm1) from the hyperthermophilic archaeon Pyrococcus furiosus:<br/>Reference Title: cloning, sequencing of the gene and its expression in<br/>Reference Title: Escherichia coli.<br/>Reference Author: Constantinesco F, Benachenhou N, Motorin Y, Grosjean H;<br/>Reference Location: Nucleic Acids Res 1998;26:3753-3761.<br/>Reference Number: [2]<br/>Reference Medline: 87260951<br/>Reference Title: Amino-terminal extension generated from an upstream AUG<br/>Reference Title: codon is not required for mitochondrial import of yeast<br/>Reference Title: N2,N2-dimethylguanosine- specific tRNA methyltransferase.<br/>Reference Author: Ellis SR, Hopper AK, Martin NC;<br/>Reference Location: Proc Natl Acad Sci U S A 1987;84:5172-5176.<br/>Database Reference: INTERPRO; IPR002905;<br/>Database reference: PFAMB; PB041661;<br/>Comment: This enzyme EC:2.1.1.32 used S-AdoMet to methylate tRNA.<br/>Comment: The TRM1 gene of Saccharomyces cerevisiae is necessary for<br/>Comment: the N2,N2-dimethylguanosine modification of both mitochondrial<br/>Comment: and cytoplasmic tRNAs [1]. The enzyme is</p> |

004007 00264960

| Pfam         | Prosite   | Full Name  | Description  |
|--------------|-----------|--|--|
|              |           |  | found in both<br>Comment: eukaryotes and archaeobacteria [2]<br>Number of members: 10  |
| tRNA_bind    |           | Putative tRNA binding domain                           | <p>Accession number: PF01588<br/> Definition: Putative tRNA binding domain<br/> Author: Bashton M, Bateman A<br/> Alignment method of seed: Clustalw<br/> Source of seed members: Pfam-B_482 (release 4.1)<br/> Gathering cutoffs: 20 20<br/> Trusted cutoffs: 22.30 22.30<br/> Noise cutoffs: 18.20 18.20<br/> HMM build command line: hmmbuild -F HMM SEED<br/> HMM build command line: hmmscalibrate --seed 0 HMM<br/> Reference Number: [1]<br/> Reference Medline: 97306356<br/> Reference Title: Human tyrosyl-tRNA synthetase shares amino acid sequence<br/> Reference Title: homology with a putative cytokine.<br/> Reference Author: Kleeman TA, Wei D, Simpson KL, First EA;<br/> Reference Location: J Biol Chem 1997;272:14420-14425.<br/> Reference Number: [2]<br/> Reference Medline: 97050848<br/> Reference Title: The yeast protein Arc1p binds to tRNA and functions as a<br/> Reference Title: cofactor for the methionyl- and glutamyl-tRNA synthetases.<br/> Reference Author: Simos G, Segref A, Fasiolo F, Hellmuth K, Shevchenko A,<br/> Reference Author: Mann M, Hurt EC;<br/> Reference Location: EMBO J 1996;15:5437-5448.<br/> Database Reference: SCOP; 1pys; fa; [SCOP-USA][CATH-PDBSUM]<br/> Database Reference: INTERPRO; IPR002547;<br/> Database Reference: PDB; 1b70 B; 153; 247;<br/> Database Reference: PDB; 1b7y B; 153; 247;<br/> Database Reference: PDB; 1eiy B; 153; 247;<br/> Database Reference: PDB; 1pys B; 153; 247;<br/> Database reference: PFAMB; PB010015;<br/> Comment: This domain is found in prokaryotic methionyl-tRNA synthetases,<br/> Comment: prokaryotic phenylalanyl tRNA synthetases<br/> the yeast GU4 nucleic-binding<br/> Comment: protein (G4p1 or p42, ARC1) [2], human tyrosyl-tRNA synthetase [1],<br/> Comment: and endothelial-monocyte activating polypeptide II.<br/> Comment: G4p1 binds specifically to tRNA form a complex with methionyl-tRNA<br/> Comment: synthetases [2]. In human tyrosyl-tRNA synthetase this domain may direct<br/> Comment: tRNA to the active site of the enzyme [2].<br/> This domain may perform a<br/> Comment: common function in tRNA aminoacylation [1].<br/> Number of members: 46</p> |
| tRNA-synt_2d | PDOC00363 | Aminoacyl-transfer RNA synthetases class-II signatures | <p>Aminoacyl-tRNA synthetases (EC 6.1.1.-) [1] are a group of enzymes which activate amino acids and transfer them to specific tRNA molecules as the first step in protein biosynthesis. In prokaryotic organisms there are at least twenty different types of aminoacyl-tRNA synthetases, one for each different amino acid. In eukaryotes there are generally two aminoacyl-tRNA synthetases for each different amino acid: one cytosolic form and a mitochondrial form. While all these enzymes have a common function, they are widely diverse in terms of subunit size and of quaternary structure.</p>   |

004001" 60267960



| Pfam    | Prosite   | Full Name                                      | Description  |
|---------|-----------|--|--|
|         |           |  | <p>The synthetases specific for alanine, asparagine, aspartic acid, glycine, histidine, lysine, phenylalanine, proline, serine, and threonine are referred to as class-II synthetases [2 to 6] and probably have a common folding pattern in their catalytic domain for the binding of ATP and amino acid which is different to the Rossmann fold observed for the class I synthetases [7].</p> <p>Class-II tRNA synthetases do not share a high degree of similarity, however at least three conserved regions are present [2,5,8]. We have derived signature patterns from two of these regions.</p> <p>Description of pattern(s) and/or profile(s)</p> <p>Consensus pattern [FYH]-R-x-[DE]-x(4,12)-[RH]-x(3)-F-x(3)-[DE]<br/>Sequences known to belong to this class detected by the pattern the majority of class-II tRNA synthetases with the exception of those specific for alanine, glycine as well as bacterial histidine. Other sequence(s) detected in SWISS-PROT 43.</p> <p>Consensus pattern [GSTALVF]-{DENQHRKP}-[GSTA]-[LIVMF]-[DE]-R-[LIVMF]-x-[LIVMSTAG]-[LIVMFY]<br/>Sequences known to belong to this class detected by the pattern the majority of class-II tRNA synthetases with the exception of those specific for serine and proline. Other sequence(s) detected in SWISS-PROT 161.</p> <p>Expert(s) to contact by email<br/>Cusack S. <a href="mailto:cusack@embl-grenoble.fr">cusack@embl-grenoble.fr</a></p> <p>Last update<br/>July 1998 / Text revised.</p> <p>References<br/>[ 1]<br/>Schimmel P.<br/>Annu. Rev. Biochem. 56:125-158(1987).</p> <p>[ 2]<br/>Delarue M., Moras D.<br/>BioEssays 15:675-687(1993).</p> <p>[ 3]<br/>Schimmel P.<br/>Trends Biochem. Sci. 16:1-3(1991).</p> <p>[ 4]<br/>Nagel G.M., Doolittle R.F.<br/>Proc. Natl. Acad. Sci. U.S.A. 88:8121-8125(1991).</p> <p>[ 5]<br/>Cusack S., Haertlein M., Leberman R.<br/>Nucleic Acids Res. 19:3489-3498(1991).</p> <p>[ 6]<br/>Cusack S.<br/>Biochimie 75:1077-1081(1993).</p> <p>[ 7]<br/>Cusack S., Berthet-Colominas C., Haertlein M., Nassar N., Leberman R.<br/>Nature 347:249-255(1990).</p> <p>[ 8]<br/>Leveque F., Plateau P., Dessen P., Blanquet S.<br/>Nucleic Acids Res. 18:305-312(1990).</p> |
| trypsin | PDOC00124 | Serine proteases, trypsin family, active sites | The catalytic activity of the serine proteases from the trypsin family is provided by a charge relay system involving an aspartic acid residue hydrogen-   |

004001" 00264960

| Pfam | Prosite | Full Name | Description   |
|------|---------|-----------|---|
|      |         |           | <p>bonded to a histidine, which itself is hydrogen-bonded to a serine. The sequences in the vicinity of the active site serine and histidine residues are well conserved in this family of proteases [1]. A partial list of proteases known to belong to the trypsin family is shown below.</p> <ul style="list-style-type: none"> <li>- Acrosin.</li> <li>- Blood coagulation factors VII, IX, X, XI and XII, thrombin, plasminogen, and protein C.</li> <li>- Cathepsin G.</li> <li>- Chymotrypsins.</li> <li>- Complement components C1r, C1s, C2, and complement factors B, D and I.</li> <li>- Complement-activating component of RA-reactive factor.</li> <li>- Cytotoxic cell proteases (granzymes A to H).</li> <li>- Duodenase I.</li> <li>- Elastases 1, 2, 3A, 3B (protease E), leukocyte (medullasin).</li> <li>- Enterokinase (EC 3.4.21.9) (enteropeptidase).</li> <li>- Hepatocyte growth factor activator.</li> <li>- Hepsin.</li> <li>- Glandular (tissue) kallikreins (including EGF-binding protein types A, B, and C, NGF-gamma chain, gamma-renin, prostate specific antigen (PSA) and tonin).</li> <li>- Plasma kallikrein.</li> <li>- Mast cell proteases (MCP) 1 (chymase) to 8.</li> <li>- Myeloblastin (proteinase 3) (Wegener's autoantigen).</li> <li>- Plasminogen activators (urokinase-type, and tissue-type).</li> <li>- Trypsins I, II, III, and IV.</li> <li>- Trypsases.</li> <li>- Snake venom proteases such as ancrod, batroxobin, cerastobin, flavoxobin, and protein C activator.</li> <li>- Collagenase from common cattle grub and collagenolytic protease from Atlantic sand fiddler crab.</li> <li>- Apolipoprotein(a).</li> <li>- Blood fluke cercarial protease.</li> <li>- Drosophila trypsin like proteases: alpha, easter, snake-locus.</li> <li>- Drosophila protease stubble (gene sb).</li> <li>- Major mite fecal allergen Der p III.</li> </ul> <p>All the above proteins belong to family S1 in the classification of peptidases [2,E1] and originate from eukaryotic species. It should be noted that bacterial proteases that belong to family S2A are similar enough in the regions of the active site residues that they can be picked up by the same patterns. These proteases are listed below.</p> <ul style="list-style-type: none"> <li>- Achromobacter lyticus protease I.</li> <li>- Lysobacter alpha-lytic protease.</li> <li>- Streptogrisin A and B (Streptomyces proteases A and B).</li> <li>- Streptomyces griseus glutamyl endopeptidase II.</li> <li>- Streptomyces fradiae proteases 1 and 2.</li> </ul> <p>Description of pattern(s) and/or profile(s)</p> <p>Consensus pattern [LIVM]-[ST]-A-[STAG]-H-C [H is the active site residue]</p> <p>Sequences known to belong to this class detected by the pattern ALL, except for complement components C1r and C1s, pig plasminogen, bovine protein C, rodent urokinase, ancrod, gyroxin and two insect trypsins.</p> <p>Other sequence(s) detected in SWISS-PROT 14.</p> |

004007 "E0264960

| Pfam       | Prosite   | Full Name              | Description  |
|------------|-----------|------------------------|--|
|            |           |                        | <p>Consensus pattern [DNSTAGC]-[GSTAPIMVQH]-x(2)-G-[DE]-S-G-[GS]-[SAPHV]-[LIVMFYWH]-[LIVMFYSTANQH] [S is the active site residue]</p> <p>Sequences known to belong to this class detected by the pattern ALL, except for 18 different proteases which have lost the first conserved glycine.</p> <p>Other sequence(s) detected in SWISS-PROT H.influenzae protease HAP which belongs to family S6 and 3 other proteins.</p> <p>Note if a protein includes both the serine and the histidine active site signatures, the probability of it being a trypsin family serine protease is 100%</p> <p>Last update<br/>November 1997 / Text revised.</p> <p>References<br/>[ 1]<br/>Brenner S.<br/>Nature 334:528-530(1988).</p> <p>[ 2]<br/>Rawlings N.D., Barrett A.J.<br/>Meth. Enzymol. 244:19-61(1994).</p> <p>[E1]<br/><a href="http://www.expasy.ch/cgi-bin/lists?peptidas.txt">http://www.expasy.ch/cgi-bin/lists?peptidas.txt</a></p>  |
| TYA        |           | TYA transposon protein | <p>Accession number: PF01021</p> <p>Definition: TYA transposon protein</p> <p>Author: Bateman A</p> <p>Alignment method of seed: Clustalw</p> <p>Source of seed members: Pfam-B_90 (release 3.0)</p> <p>Gathering cutoffs: 15 15</p> <p>Trusted cutoffs: 18.00 18.00</p> <p>Noise cutoffs: 13.70 13.70</p> <p>HMM build command line: hmmbuild -f HMM SEED</p> <p>HMM build command line: hmmscalibrate --seed 0 HMM</p> <p>Reference Number: [1]</p> <p>Reference Medline: 97404699</p> <p>Reference Title: Cryo-electron microscopy structure of yeast Ty</p> <p>Reference Title: retrotransposon virus-like particles.</p> <p>Reference Author: Palmer KJ, Tichelaar W, Myers N, Burns NR, Butcher SJ,</p> <p>Reference Author: Kingsman AJ, Fuller SD, Saibil HR;</p> <p>Reference Location: J Virol 1997;71:6863-6868.</p> <p>Database Reference: INTERPRO; IPR001042;</p> <p>Comment: Ty are yeast transposons. A 5.7kb transcript codes</p> <p>Comment: for p3 a fusion protein of TYA and TYB.</p> <p>The TYA</p> <p>Comment: protein is analogous to the gag protein of retroviruses.</p> <p>Comment: TYA a is cleaved to form 46kd protein which can form</p> <p>Comment: mature virion like particles [1].</p> <p>Number of members: 62</p> |
| tyrosinase | PDOC00398 | Tyrosinase signatures  | <p>Tyrosinase (EC 1.14.18.1) [1] is a copper monooxygenases that catalyzes the hydroxylation of monophenols and the oxidation of o-diphenols to o-quinols.</p> <p>This enzyme, found in prokaryotes as well as in eukaryotes, is involved in the formation of pigments such as melanins and other polyphenolic compounds.</p> <p>Tyrosinase binds two copper ions (CuA and CuB). Each of the two copper ion has been shown [2] to be bound by three conserved histidines residues. The regions around these copper-binding ligands are well conserved and also shared by some hemocyanins, which are copper-containing oxygen carriers from the hemolymph of</p>   |

09679203 100400

| Pfam | Prosite | Full Name | Description  |
|------|---------|-----------|--|
|      |         |           | <p>many molluscs and arthropods [3,4].</p> <p>At least two proteins related to tyrosinase are known to exist in mammals:</p> <ul style="list-style-type: none"> <li>- TRP-1 (TYRP1) [5], which is responsible for the conversion of 5,6-dihydroxyindole-2-carboxylic acid (DHICA) to indole-5,6-quinone-2-carboxylic acid.</li> <li>- TRP-2 (TYRP2) [6], which is the melanogenic enzyme DOPachrome tautomerase (EC 5.3.3.12) that catalyzes the conversion of DOPachrome to DHICA. TRP-2 differs from tyrosinases and TRP-1 in that it binds two zinc ions instead of copper [7].</li> </ul> <p>Other proteins that belong to this family are:</p> <ul style="list-style-type: none"> <li>- Plants polyphenol oxidases (PPO) (EC 1.10.3.1) which catalyze the oxidation of mono- and o-diphenols to o-diquinones [8].</li> <li>- Caenorhabditis elegans hypothetical protein C02C2.1.</li> </ul> <p>We have derived two signature patterns for tyrosinase and related proteins. The first one contains two of the histidines that bind CuA, and is located in the N-terminal section of tyrosinase. The second pattern contains a histidine that binds CuB, that pattern is located in the central section of the enzyme.</p> <p>Description of pattern(s) and/or profile(s)</p> <p>Consensus pattern H-x(4,5)-F-[LVMFTP]-x-[FW]-H-R-x(2)-[LVM]-x(3)-E [The two H's are copper ligands]<br/>Sequences known to belong to this class detected by the pattern ALL.<br/>Other sequence(s) detected in SWISS-PROT NONE.</p> <p>Consensus pattern D-P-x-F-[LVMFYW]-x(2)-H-x(3)-D [H is a copper ligand]<br/>Sequences known to belong to this class detected by the pattern ALL the tyrosinases as well as all the hemocyanins.<br/>Other sequence(s) detected in SWISS-PROT NONE.<br/>Last update<br/>December 1999 / Patterns and text revised.<br/>References<br/>[ 1]<br/>Lerch K.<br/>Prog. Clin. Biol. Res. 256:85-98(1988).<br/>[ 2]<br/>Jackman M.P., Hajnal A., Lerch K.<br/>Biochem. J. 274:707-713(1991).<br/>[ 3]<br/>Linzen B.<br/>Naturwissenschaften 76:206-211(1989).<br/>[ 4]<br/>Lang W.H., van Holde K.E.<br/>Proc. Natl. Acad. Sci. U.S.A. 88:244-248(1991).<br/>[ 5]<br/>Kobayashi T., Urabe K., Winder A., Jimenez-Cervantes C., Imokawa G., Brewington T., Solano F., Garcia-Borrón J.C., Hearing V.J.<br/>EMBO J. 13:5818-5825(1994).<br/>[ 6]</p> |

09679203 100400

| Pfam            | Prosite   | Full Name                                     | Description   |
|-----------------|-----------|---|---|
|                 |           |   | <p>Jackson I.J., Chambers D.M., Tsukamoto K., Copeland N.G., Gilbert D.J., Jenkins N.A., Hearing V.<br/>EMBO J. 11:527-535(1992).</p> <p>[ 7 ]<br/>Solano F., Martinez-Liarte J.H., Jimenez-Cervantes C., Garcia-Borron J.C., Lozano J.A.<br/>Biochem. Biophys. Res. Commun. 204:1243-1250(1994).</p> <p>[ 8 ]<br/>Cary J.W., Lax A.R., Flurkey W.H.<br/>Plant Mol. Biol. 20:245-253(1992).</p>   |
| UbiA            | PDOC00727 | UbiA prenyltransferase family signature       | <p>The following prenyltransferases are evolutionary related [1,2]:</p> <ul style="list-style-type: none"> <li>- Bacterial 4-hydroxybenzoate octaprenyltransferase (gene ubiA).</li> <li>- Yeast mitochondrial para-hydroxybenzoate--polyprenyltransferase (gene COQ2).</li> <li>- Protoheme IX farnesyltransferase (heme O synthase) from yeast and mammals (gene COX10) and from bacteria (genes cyoE or ctaB).</li> </ul> <p>These proteins probably contain seven transmembrane segments. The best conserved region is located in a loop between the second and third of these segments and we used it as a signature pattern.</p> <p>Description of pattern(s) and/or profile(s)</p> <p>Consensus pattern N-x(3)-[DEH]-x(2)-[LIMF]-D-x(2)-[VM]-x-R-[ST]-x(2)-R-x(4)-G<br/>Sequences known to belong to this class detected by the pattern ALL.<br/>Other sequence(s) detected in SWISS-PROT NONE.<br/>Last update<br/>December 1999 / Pattern and text revised.<br/>References<br/>[ 1 ]<br/>Melzer M., Heide L.<br/>Biochim. Biophys. Acta 1212:93-102(1994).</p> <p>[ 2 ]<br/>Mogi T., Saiki K., Anraku Y.<br/>Mol. Microbiol. 14:391-398(1994).</p> |
| Ubie_methyltran | PDOC00911 | ubiE/COQ5 methyltransferase family signatures | <p>The following methyltransferases have been shown [1] to share regions of similarities:</p> <ul style="list-style-type: none"> <li>- Escherichia coli ubiE, which is involved in both ubiquinone and menaquinone biosynthesis and which catalyzes the S-adenosylmethionine dependent methylation of 2-polyprenyl-6-methoxy-1,4-benzoquinol into 2-polyprenyl-3-methyl-6-methoxy-1,4-benzoquinol and of demethylmenaquinol into menaquinol.</li> <li>- Yeast COQ5, a ubiquinone biosynthesis methyltransferase.</li> <li>- Bacillus subtilis spore germination protein C2 (gene: gercB or gerC2), a probable menaquinone biosynthesis methyltransferase.</li> <li>- Lactococcus lactis gerC2 homolog.</li> <li>- Caenorhabditis elegans hypothetical protein ZK652.9.</li> <li>- Leishmania donovani amastigote-specific protein A41.</li> </ul> <p>These are hydrophilic proteins of about 30 Kd (except for ZK652.9 which is 65 Kd). They can be picked up in the database by the following patterns.</p>  |

004001 00064960

| Pfam      | Prosite   | Full Name                              | Description   |
|-----------|-----------|--|---|
|           |           |  | <p>Description of pattern(s) and/or profile(s)</p> <p>Consensus pattern Y-D-x-M-N-x(2)-[LIVM]-S-x(3)-H-x(2)-W<br/>Sequences known to belong to this class detected by the pattern ALL.<br/>Other sequence(s) detected in SWISS-PROT NONE.</p> <p>Consensus pattern R-V-[LIVM]-K-[PV]-[GM]-G-x-[LIVMF]-x(2)-[LIVM]-E-x-S<br/>Sequences known to belong to this class detected by the pattern ALL.<br/>Other sequence(s) detected in SWISS-PROT NONE.</p> <p>Last update<br/>December 1999 / Pattern and text revised.<br/>References<br/>[ 1]<br/>Lee P.T., Hsu A.Y., Ha H.T., Clarke C.F.<br/>J. Bacteriol. 179:1748-1754(1997).</p>  |
| ubiquitin | PDOC00271 | Ubiquitin domain signature and profile | <p>Ubiquitin [1,2,3] is a protein of seventy six amino acid residues, found in all eukaryotic cells and whose sequence is extremely well conserved from protozoan to vertebrates. It plays a key role in a variety of cellular processes, such as ATP-dependent selective degradation of cellular proteins, maintenance of chromatin structure, regulation of gene expression, stress response and ribosome biogenesis.</p> <p>In most species, there are many genes coding for ubiquitin. However they can be classified into two classes. The first class produces polyubiquitin molecules consisting of exact head to tail repeats of ubiquitin. The number of repeats is variable (up to twelve in a <i>Xenopus</i> gene). In the majority of polyubiquitin precursors, there is a final amino-acid after the last repeat.</p> <p>The second class of genes produces precursor proteins consisting of a single copy of ubiquitin fused to a C-terminal extension protein (CEP). There are two types of CEP proteins and both seem to be ribosomal proteins.</p> <p>Ubiquitin is a globular protein, the last four C-terminal residues (Leu-Arg-Gly-Gly) extending from the compact structure to form a 'tail', important for its function. The latter is mediated by the covalent conjugation of ubiquitin to target proteins, by an isopeptide linkage between the C-terminal glycine and the epsilon amino group of lysine residues in the target proteins.</p> <p>There are a number of proteins which are evolutionary related to ubiquitin:</p> <ul style="list-style-type: none"> <li>- Ubiquitin-like proteins from baculoviruses as well as in some strains of bovine viral diarrhea viruses (BVDV). These proteins are highly similar to their eukaryotic counterparts.</li> <li>- Mammalian protein GDX [4]. GDX is composed of two domains, a N-terminal ubiquitin-like domain of 74 residues and a C-terminal domain of 83 residues with some similarity with the thyroglobulin hormonogenic site.</li> <li>- Mammalian protein FAU [5]. FAU is a fusion protein which</li> </ul> |

004001 " E026/950

| Pfam | Prosite | Full Name | Description  |
|------|---------|-----------|--|
|      |         |           | <p>consist of a</p> <ul style="list-style-type: none"> <li>- N-terminal ubiquitin-like protein of 74 residues fused to ribosomal protein S30.</li> <li>- Mouse protein NEDD-8 [6], a ubiquitin-like protein of 81 residues.</li> <li>- Human protein BAT3, a large fusion protein of 1132 residues that contains a</li> <li>- N-terminal ubiquitin-like domain.</li> <li>- Caenorhabditis elegans protein ubl-1 [7]. Ubl-1 is a fusion protein which</li> <li>- consist of a N-terminal ubiquitin-like protein of 70 residues fused to</li> <li>- ribosomal protein S27A.</li> <li>- Yeast DNA repair protein RAD23 [8]. RAD23 contains a N-terminal domain that</li> <li>- seems to be distantly, yet significantly, related to ubiquitin.</li> <li>- Mammalian RAD23-related proteins RAD23A and RAD23B.</li> <li>- Mammalian BCL-2 binding athanogene-1 (BAG-1). BAG-1 is a protein of 274</li> <li>- residues that contains a central ubiquitin-like domain.</li> <li>- Human spliceosome associated protein 114 (SAP 114 or SF3A120).</li> <li>- Yeast protein DSK2, a protein involved in spindle pole body duplication and</li> <li>- which contains a N-terminal ubiquitin-like domain.</li> <li>- Human protein CKAP1/TFCB, Schizosaccharomyces pombe protein alp11 and</li> <li>- Caenorhabditis elegans hypothetical protein F53F4.3. These proteins contain</li> <li>- a N-terminal ubiquitin domain and a C-terminal CAP-Gly domain (see &lt;PDOC00660&gt;).</li> <li>- Schizosaccharomyces pombe hypothetical protein SpAC26A3.16. This protein</li> <li>- contains a N-terminal ubiquitin domain.</li> <li>- Yeast protein SMT3.</li> <li>- Human ubiquitin-like proteins SMT3A and SMT3B.</li> <li>- Human ubiquitin-like protein SMT3C (also known as PIC1; Ubl1, Sumo-1; Gmp-1</li> <li>- or Sentrin). This protein is involved in targeting ranGAP1 to the nuclear</li> <li>- pore complex protein ranBP2.</li> <li>- SMT3-like proteins in plants and Caenorhabditis elegans.</li> </ul> <p>To identify ubiquitin and related proteins we have developed a pattern based on conserved positions in the central section of the sequence. A profile was also developed that spans the complete length of the ubiquitin domain.</p> <p>Description of pattern(s) and/or profile(s)</p> <p>Consensus pattern K-x(2)-[LIVM]-x-[DESAK]-x(3)-[LIVM]-[PA]-x(3)-Q-x-[LIVM]-[LIVMC]-[LIVMFY]-x-G-x(4)-[DE]</p> <p>Sequences known to belong to this class detected by the pattern ALL, except for the RAD23 and SMT3 subfamilies, BAG-1 and SAP 114.</p> <p>Other sequence(s) detected in SWISS-PROT NONE.</p> <p>Sequences known to belong to this class detected by the profile ALL.</p> <p>Other sequence(s) detected in SWISS-PROT NONE.</p> <p>Note this documentation entry is linked to both a signature pattern and a profile. As the profile is much more sensitive than the pattern, you should use it if you have access to the necessary software tools to do so.</p> <p>Last update<br/>July 1998 / Text revised.</p> |

004001" E0254960

| Pfam    | Prosite   | Full Name  | Description   |
|---------|-----------|--|---|
|         |           |  | <p>Bio/Technology 8:209-215(1990). References</p> <p>[ 1 ]<br/>Jentsch S., Seufert W., Hauser H.-P.<br/>Biochim. Biophys. Acta 1089:127-139(1991).</p> <p>[ 2 ]<br/>Monia B.P., Ecker D.J., Croke S.T.</p> <p>[ 3 ]<br/>Finley D., Varshavsky A.<br/>Trends Biochem. Sci. 10:343-347(1985).</p> <p>[ 4 ]<br/>Filippi M., Tribioli C., Toniolo D.<br/>Genomics 7:453-457(1990).</p> <p>[ 5 ]<br/>Olvera J., Wool I.G.<br/>J. Biol. Chem. 268:17967-17974(1993).</p> <p>[ 6 ]<br/>Kumar S., Yoshida Y., Noda M.<br/>Biochem. Biophys. Res. Commun. 195:393-399(1993).</p> <p>[ 7 ]<br/>Jones D., Candido E.P.<br/>J. Biol. Chem. 268:19545-19551(1993).</p> <p>[ 8 ]<br/>Melnick L., Sherman F.<br/>J. Mol. Biol. 233:372-388(1993).</p>  |
| UPF0004 | PDOC00984 | Uncharacterized protein family UPF0004 signature | <p>The following uncharacterized proteins have been shown [1] to share regions of similarities:</p> <ul style="list-style-type: none"> <li>- Escherichia coli hypothetical protein yliG.</li> <li>- Escherichia coli hypothetical protein yleA and HI0019, the corresponding Haemophilus influenzae protein.</li> <li>- Bacillus subtilis hypothetical protein yqeV.</li> <li>- Helicobacter pylori hypothetical protein HP0269.</li> <li>- Helicobacter pylori hypothetical protein HP0285.</li> <li>- Mycoplasma iowae hypothetical protein in 16S RNA 5' region.</li> <li>- Mycobacterium tuberculosis hypothetical protein Rv2733c.</li> <li>- Rickettsia prowazekii hypothetical protein RP416.</li> <li>- Rickettsia prowazekii hypothetical protein RP808.</li> <li>- Synechocystis strain PCC 6803 hypothetical protein slr0082.</li> <li>- Synechocystis strain PCC 6803 hypothetical protein slr0996.</li> <li>- Methanococcus jannaschii hypothetical protein MJ0865.</li> <li>- Methanococcus jannaschii hypothetical protein MJ0867.</li> <li>- Caenorhabditis elegans hypothetical protein F25B5.5.</li> </ul> <p>The size of these proteins range from 47 to 61 Kd. They contain six conserved cysteines, three of which are clustered in a region that can be used as a signature pattern.</p> <p>Description of pattern(s) and/or profile(s)</p> <p>Consensus pattern [LIVM]-x-[LIVMT]-x(2)-G-C-x(3)-C-[STAN]-[FY]-C-x-[LIVMT]-x(4)-G</p> <p>Sequences known to belong to this class detected by the pattern ALL.</p> <p>Other sequence(s) detected in SWISS-PROT 2.</p> <p>Last update<br/>December 1999 / Pattern and text revised.</p> <p>References<br/>[ 1 ]<br/>Bairoch A.</p> |

004001" E026/960



| Pfam    | Prosite   | Full Name  | Description   |
|---------|-----------|--|---|
|         |           |  | Unpublished observations (1997).  |
| UPF0013 |           | Uncharacterized membrane protein family UPF0013  | <p>Accession number: PF01554</p> <p>Definition: Uncharacterized membrane protein family UPF0013</p> <p>Author: Bateman A</p> <p>Alignment method of seed: Clustalw</p> <p>Source of seed members: Pfam-B_163 (release 4.0)</p> <p>Gathering cutoffs: -26 -26</p> <p>Trusted cutoffs: -16.10 -16.10</p> <p>Noise cutoffs: -36.70 -36.70</p> <p>HMM build command line: hmmbuild -F HMM SEED</p> <p>HMM build command line: hmmcalibrate --seed 0 HMM</p> <p>Database Reference: URL; <a href="http://www.expasy.ch/cgi-bin/lists?upflist.txt">http://www.expasy.ch/cgi-bin/lists?upflist.txt</a>;</p> <p>Database Reference: INTERPRO; IPR002528;</p> <p>Database reference: PFAMB; PB041103;</p> <p>Comment: These proteins are integral membrane proteins of unknown</p> <p>Comment: function.</p> <p>Number of members: 47</p>  |
| UPF0019 | PDOC00949 | Uncharacterized protein family UPF0019 signature | <p>The following uncharacterized proteins have been shown [1,2] to be highly similar:</p> <ul style="list-style-type: none"> <li>- Yeast protein SNZ1, which may be involved in growth arrest and cellular response to nutrient limitation.</li> <li>- Yeast chromosome VI hypothetical protein YFL059w.</li> <li>- Yeast chromosome XIV hypothetical protein YNL333w.</li> <li>- Fission yeast hypothetical protein SpAC29B12.04.</li> <li>- Hevea brasiliensis ethylene-inducible protein HEVER.</li> <li>- Stellaria longipes hypothetical protein H47.</li> <li>- Bacillus subtilis hypothetical protein yaaD.</li> <li>- Haemophilus influenzae hypothetical protein HI1647.</li> <li>- Mycobacterium leprae hypothetical protein M1CL581.12c.</li> <li>- Mycobacterium tuberculosis hypothetical protein MtCY1A10.27.</li> <li>- Archaeoglobus fulgidus hypothetical protein AF0508.</li> <li>- Methanococcus jannaschii hypothetical protein MJ0677.</li> <li>- Methanococcus vannielii hypothetical protein in tRNA/5S rRNA gene cluster.</li> <li>- Methanobacterium thermoautotrophicum hypothetical protein Mth666.</li> </ul> <p>These are hydrophilic proteins of about 32 Kd. They can be picked up in the database by the following pattern.</p> <p>Description of pattern(s) and/or profile(s)</p> <p>Consensus pattern L-P-V-[VT]-[NQL]-F-[AT]-A-G-G-[LIV]-A-T-P-A-D-A-A-[LM]</p> <p>Sequences known to belong to this class detected by the pattern ALL.</p> <p>Other sequence(s) detected in SWISS-PROT NONE.</p> <p>Last update</p> <p>July 1998 / Pattern and text revised.</p> <p>References</p> <p>[1]</p> <p>Sivasubramaniam S., Vanniasingham V.M., Tan C.T., Chua N.H. Plant Mol. Biol. 29:173-178(1995).</p> <p>[2]</p> <p>Braun E.L., Fuge E.K., Padilla P.A., Werner-Washburne M. J. Bacteriol. 178:6865-6872(1996).</p> |
| UPF0047 | PDOC01018 | Uncharacterized protein family UPF0047 signature | <p>The following uncharacterized proteins have been shown [1] to be highly similar:</p> <ul style="list-style-type: none"> <li>- Bacillus subtilis hypothetical protein yugU.</li> </ul>  |

00679203 100400

| Pfam    | Prosite   | Full Name  | Description  |
|---------|-----------|--|--|
|         |           |  | <ul style="list-style-type: none"> <li>- Escherichia coli hypothetical protein yjbQ.</li> <li>- Mycobacterium tuberculosis hypothetical protein MtCY9C4.12.</li> <li>- Synechocystis strain PCC 6803 hypothetical protein sli1880.</li> <li>- Archaeoglobus fulgidus hypothetical protein AF2050.</li> <li>- Methanococcus jannaschii hypothetical protein MJ1081.</li> <li>- Methanobacterium thermoautotrophicum hypothetical protein MTH771.</li> <li>- Fission yeast hypothetical protein SpAC4A8.02c.</li> </ul> <p>These are small proteins of 14 to 16 Kd. They can be picked up in the database by the following pattern. This pattern is located in the C-terminal part of these proteins.</p> <p>Description of pattern(s) and/or profile(s)</p> <p>Consensus pattern S-X(2)-[LIV]-x-[LIV]-x(2)-G-x(4)-G-T-W-Q-x-[LIV]</p> <p>Sequences known to belong to this class detected by the pattern ALL.</p> <p>Other sequence(s) detected in SWISS-PROT NONE.</p> <p>Last update July 1998 / First entry.</p> <p>References [1]</p> <p>Bairoch A.</p> <p>Unpublished observations (1998).</p>       |
| UPF0052 |           | Uncharacterised protein family UPF0052           | <p>Accession number: PF01933</p> <p>Definition: Uncharacterised protein family UPF0052</p> <p>Author: Enright A, Ouzounis C, Bateman A</p> <p>Alignment method of seed: Clustalw</p> <p>Source of seed members: Enright A</p> <p>Gathering cutoffs: 25 25</p> <p>Trusted cutoffs: 263.90 263.90</p> <p>Noise cutoffs: -134.40 -134.40</p> <p>HMM build command line: hmmbuild -F HMM SEED</p> <p>HMM build command line: hmmcalibrate --seed 0 HMM</p> <p>Database Reference INTERPRO; IPR002882;</p> <p>Number of members: 12</p>   |
| UPF0057 | PDOC01013 | Uncharacterized protein family UPF0057 signature | <p>The following uncharacterized proteins have been shown [1] to be evolutionary related:</p> <ul style="list-style-type: none"> <li>- Barley low-temperature induced protein blt101.</li> <li>- Lophorium elongatum salt-stress induced protein ESI3.</li> <li>- Yeast hypothetical proteins YDL123w, YDR276c, YDR525Bw and YJL151c.</li> <li>- Caenorhabditis elegans hypothetical proteins F47B7.1, T23F2.3, T23F2.4, T23F2.5 and ZK632.10.</li> <li>- Escherichia coli hypothetical protein yqaE.</li> <li>- Synechocystis strain PCC 6803 hypothetical protein ssr1169.</li> </ul> <p>These are small proteins of from 52 to 140 amino-acid residues that contains two transmembrane domains. As a signature pattern we selected a region that corresponds to the end of the first transmembrane helix.</p> <p>Description of pattern(s) and/or profile(s)</p> <p>Consensus pattern [LIV]-x-[STA]-[LIVF](3)-P-P-[LIV]-[GA]-[IV]-x(4)-[GKN]</p> <p>Sequences known to belong to this class detected by the pattern ALL.</p> <p>Other sequence(s) detected in SWISS-PROT NONE.</p> <p>Last update</p> |

004001" E026'960

| Pfam    | Prosite   | Full Name  | Description   |
|---------|-----------|--|---|
|         |           |  | <p>July 1998 / First entry.</p> <p>References</p> <p>[ 1 ]</p> <p>Rudd K.E., Humphery-Smith I., Wasinger V.C., Bairoch A. Electrophoresis 19:536-544(1998).</p>   |
| UPF0066 | PDOC01022 | Uncharacterized protein family UPF0066 signature | <p>The following uncharacterized proteins have been shown [1] to be evolutionary related:</p> <ul style="list-style-type: none"> <li>- Escherichia coli hypothetical protein yaeB and HI0510, the corresponding Haemophilus influenzae protein.</li> <li>- Agrobacterium tumefaciens Ti plasmid protein virR.</li> <li>- Pseudomonas aeruginosa protein rcsF.</li> <li>- Archaeoglobus fulgidus hypothetical protein AF0241.</li> <li>- Archaeoglobus fulgidus hypothetical protein AF0433.</li> <li>- Methanococcus jannaschii hypothetical protein MJ1583.</li> <li>- Methanobacterium thermoautotrophicum hypothetical protein MTH1797.</li> </ul> <p>These are proteins of from 120 to 240 amino-acid residues (with the exception of AF0433 which is 366 residues long). As a signature pattern we selected a conserved region in the central part of these proteins.</p> <p>Description of pattern(s) and/or profile(s)</p> <p>Consensus pattern G-[AV]-F-[STA]-x-R-[SA]-x(2)-R-P-N</p> <p>Sequences known to belong to this class detected by the pattern ALL.</p> <p>Other sequence(s) detected in SWISS-PROT NONE.</p> <p>Last update</p> <p>July 1999 / First entry.</p> <p>References</p> <p>[ 1 ]</p> <p>Bairoch A.</p> <p>Unpublished observations (1998).</p>   |
| UPF0076 | PDOC00838 | Uncharacterized protein family UPF0076 signature | <p>The following uncharacterized proteins have been shown [1] to share regions of similarities:</p> <ul style="list-style-type: none"> <li>- Goat antigen UK114, a human homolog and the rat corresponding protein which is known as perchloric acid soluble protein (PSP1). PSP1 [2] may inhibit an initiation stage of cell-free protein synthesis.</li> <li>- Mouse heat-responsive protein HRSP12.</li> <li>- Yeast chromosome V hypothetical protein YER057c.</li> <li>- Yeast chromosome IX hypothetical protein YIL051c.</li> <li>- Caenorhabditis elegans hypothetical protein C23G10.2.</li> <li>- Escherichia coli hypothetical protein ycdK.</li> <li>- Escherichia coli hypothetical protein yhaR.</li> <li>- Escherichia coli hypothetical protein yjgF and HI0719, the corresponding Haemophilus influenzae protein.</li> <li>- Escherichia coli hypothetical protein yoaB.</li> <li>- Bacillus subtilis hypothetical protein yabJ.</li> <li>- Haemophilus influenzae hypothetical protein HI1627.</li> <li>- Helicobacter pylori hypothetical protein HP0944.</li> <li>- Lactococcus lactis aldR.</li> <li>- Myxococcus xanthus dfrA.</li> <li>- Synechocystis strain PCC 6803 hypothetical protein slr0709.</li> <li>- Rhizobium strain NGR234 symbiotic plasmid hypothetical protein y4sK.</li> <li>- Pyrococcus horikoshii hypothetical protein PH0854.</li> </ul> <p>These are small proteins of around 15 Kd whose sequence is highly conserved.</p> <p>As a signature pattern, we selected a well conserved region</p> |

004007 "E0267960

| Pfam    | Prosite   | Full Name                                 | Description  |
|---------|-----------|---|--|
|         |           |   | <p>located in the C-terminal part of these proteins.</p> <p>Description of pattern(s) and/or profile(s)</p> <p>Consensus pattern [PA]-[ASTPV]-R-[SACVF]-x-[LIVMFY]-x(2)-[GSAKR]-x-[LMVA]-x(5,8)-[LIVM]-E-[MI]</p> <p>Sequences known to belong to this class detected by the pattern ALL.</p> <p>Other sequence(s) detected in SWISS-PROT 4.</p> <p>Last update<br/>July 1999 / Pattern and text revised.</p> <p>References<br/>[ 1]<br/>Bairoch A.<br/>Unpublished observations (1995).</p> <p>[ 2]<br/>Oka T., Tsuji H., Noda C., Sakai K., Hong Y.-M., Suzuki I., Munoz S., Natori Y.<br/>J. Biol. Chem. 270:30060-30067(1995).</p>   |
| UPF0099 |           | Domain of unknown function UPF0099        | <p>Accession number: PF01981</p> <p>Definition: Domain of unknown function UPF0099</p> <p>Previous Pfam IDs: DUF119;</p> <p>Author: Enright A, Ouzounis C, Bateman A</p> <p>Alignment method of seed: Clustalw</p> <p>Source of seed members: Enright A</p> <p>Gathering cutoffs: 25 25</p> <p>Trusted cutoffs: 132.80 132.80</p> <p>Noise cutoffs: -35.70 -35.70</p> <p>HMM build command line: hmmbuild -F HMM SEED</p> <p>HMM build command line: hmmscalibrate --seed 0 HMM</p> <p>Database Reference INTERPRO; IPR002833;</p> <p>Comment: This domain has no known function.</p> <p>Number of members: 10</p>   |
| UQ_con  | PDOC00163 | Ubiquitin-conjugating enzymes active site | <p>Ubiquitin-conjugating enzymes (EC 6.3.2.19) (UBC or E2 enzymes) [1,2,3] catalyze the covalent attachment of ubiquitin to target proteins. An activated ubiquitin moiety is transferred from an ubiquitin-activating enzyme (E1) to E2 which later ligates ubiquitin directly to substrate proteins with or without the assistance of 'N-end' recognizing proteins (E3).</p> <p>In most species there are many forms of UBC (at least 9 in yeast) which are implicated in diverse cellular functions.</p> <p>A cysteine residue is required for ubiquitin-thiolester formation. There is a single conserved cysteine in UBC's and the region around that residue is conserved in the sequence of known UBC isozymes. We have used that region as a signature pattern.</p> <p>Description of pattern(s) and/or profile(s)</p> <p>Consensus pattern [FYWLSP]-H-[PC]-[NH]-[LIV]-x(3,4)-G-x-[LIV]-C-[LIV]-x-[LIV] [C is the active site residue]</p> <p>Sequences known to belong to this class detected by the pattern ALL, except for yeast UBC6 (DOA2).</p> <p>Other sequence(s) detected in SWISS-PROT NONE.</p> <p>Expert(s) to contact by email<br/>Jentsch S. jentsch@zmbh.uni-heidelberg.de</p> <p>Last update</p> |

| Pfam         | Prosite   | Full Name                     | Description   |
|--------------|-----------|-------------------------------|---|
|              |           |                               | <p>July 1998 / Text revised.</p> <p>References</p> <p>[ 1 ]<br/>Jentsch S., Seufert W., Sommer T., Reins H.-A.<br/>Trends Biochem. Sci. 15:195-198(1990).</p> <p>[ 2 ]<br/>Jentsch S., Seufert W., Hauser H.-P.<br/>Biochim. Biophys. Acta 1089:127-139(1991).</p> <p>[ 3 ]<br/>Hershko A.<br/>Trends Biochem. Sci. 16:265-268(1991).</p>   |
| urease_gamma | PDOC00133 | Urease signatures             | <p>Urease (EC 3.5.1.5) is a nickel-binding enzyme that catalyzes the hydrolysis of urea to carbon dioxide and ammonia [1]. Historically, it was the first enzyme to be crystallized (in 1926). It is mainly found in plant seeds, microorganisms and invertebrates. In plants, urease is a hexamer of identical chains. In bacteria [2], it consists of either two or three different subunits (alpha, beta and gamma).</p> <p>Urease binds two nickel ions per subunit; four histidine, an aspartate and a carbamated-lysine serve as ligands to these metals; an additional histidine is involved in the catalytic mechanism [3].</p> <p>As signatures for this enzyme, we selected a region that contains two histidine that bind one of the nickel ions and the region of the active site histidine.</p> <p>Description of pattern(s) and/or profile(s)</p> <p>Consensus pattern T-[AY]-[GA]-[GAT]-[LIVM]-D-x-H-[LIVM]-H-x(3)-P [The two H's bind nickel]<br/>Sequences known to belong to this class detected by the pattern ALL.<br/>Other sequence(s) detected in SWISS-PROT NONE.</p> <p>Consensus pattern [LIVM](2)-[CT]-H-[HN]-L-x(3)-[LIVM]-x(2)-D-[LIVM]-x-F-A [H is the active site residue]<br/>Sequences known to belong to this class detected by the pattern ALL.<br/>Other sequence(s) detected in SWISS-PROT NONE.</p> <p>Last update<br/>November 1997 / Patterns and text revised.</p> <p>References</p> <p>[ 1 ]<br/>Takishima K., Suga T., Mamiya G.<br/>Eur. J. Biochem. 175:151-165(1988).</p> <p>[ 2 ]<br/>Mobley H.L.T., Husinger R.P.<br/>Microbiol. Rev. 53:85-108(1989).</p> <p>[ 3 ]<br/>Jabri E., Carr M.B., Hausinger R.P., Karplus P.A.<br/>Science 268:998-1004(1995).</p> |
| UreD         |           | UreD urease accessory protein | <p>Accession number: PF01774</p> <p>Definition: UreD urease accessory protein</p> <p>Author: Bashton M, Bateman A</p> <p>Alignment method of seed: Clustalw</p> <p>Source of seed members: Pfam-B_1109 (release 4.2)</p> <p>Gathering cutoffs: 25 25</p>  |

004007 02264960

| Pfam | Prosite | Full Name | Description   |
|------|---------|-----------|---|
|      |         |           | <p>Trusted cutoffs: 186.00 186.00<br/> Noise cutoffs: -42.60 -42.60<br/> HMM build command line: hmmbuild -F HMM SEED<br/> HMM build command line: hmmscalibrate --seed 0 HMM<br/> Reference Number: [1]<br/> Reference Medline: 97352660<br/> Reference Title: Characterization of UreG, identification of a<br/> Reference Title: UreD-UreF-UreG complex, and evidence<br/> suggesting that a<br/> Reference Title: nucleotide-binding site in UreG is required<br/> for in vivo<br/> Reference Title: metallocenter assembly of Klebsiella<br/> aerogenes urease.<br/> Reference Author: Moncrief MB, Hausinger RP;<br/> Reference Location: J Bacteriol 1997;179:4081-4086.<br/> Reference Number: [2]<br/> Reference Medline: 96146510<br/> Reference Title: Organization of Ureaplasma urealyticum<br/> urease gene cluster<br/> Reference Title: and expression in a suppressor strain of<br/> Escherichia coli.<br/> Reference Author: Neyrolles O, Ferris S, Behbahani N,<br/> Montagnier L, Blanchard<br/> Reference Author: A;<br/> Reference Location: J Bacteriol 1996;178:647-655.<br/> Reference Number: [3]<br/> Reference Medline: 94211837<br/> Reference Title: In vitro activation of urease apoprotein and<br/> role of UreD<br/> Reference Title: as a chaperone required for nickel<br/> metallocenter assembly.<br/> Reference Author: Park IS, Carr MB, Hausinger RP;<br/> Reference Location: Proc Natl Acad Sci U S A 1994;91:3233-<br/> 3237.<br/> Database Reference INTERPRO; IPR002669;<br/> Comment: UreD is a urease accessory protein. Urease<br/> urease hydrolyses<br/> Comment: urea into ammonia and carbamic acid [2].<br/> UreD is involved in<br/> Comment: activation of the urease enzyme via the<br/> UreD-UreF-UreG-urease complex<br/> Comment: [1] and is required for urease nickel<br/> metallocenter assembly [3].<br/> Comment: See also UreF UreF, UreG HypB_UreG.<br/> Number of members: 23</p> |
| UreF |         | UreF      | <p>Accession number: PF01730<br/> Definition: UreF<br/> Author: Bashton M, Bateman A<br/> Alignment method of seed: Clustalw<br/> Source of seed members: Pfam-B_2037 (release 4.1)<br/> Gathering cutoffs: -31 -31<br/> Trusted cutoffs: -14.30 -14.30<br/> Noise cutoffs: -49.30 -49.30<br/> HMM build command line: hmmbuild -F HMM SEED<br/> HMM build command line: hmmscalibrate --seed 0 HMM<br/> Reference Number: [1]<br/> Reference Medline: 96404789<br/> Reference Title: Purification and activation properties of<br/> UreD-UreF-urease<br/> Reference Title: apoprotein complexes.<br/> Reference Author: Moncrief MB, Hausinger RP;<br/> Reference Location: J Bacteriol 1996;178:5417-5421.<br/> Reference Number: [2]<br/> Reference Medline: 96146510<br/> Reference Title: Organization of Ureaplasma urealyticum<br/> urease gene cluster<br/> Reference Title: and expression in a suppressor strain of<br/> Escherichia coli.<br/> Reference Author: Neyrolles O, Ferris S, Behbahani N,<br/> Montagnier L, Blanchard<br/> Reference Author: A;<br/> Reference Location: J Bacteriol 1996;178:647-655.<br/> Database Reference INTERPRO; IPR002639;</p>  |

09679203-100400

| Pfam | Prosite | Full Name                                  | Description   |
|------|---------|--|---|
|      |         |  | <p>Comment: This family consists of the Urease accessory protein</p> <p>Comment: UreF. The urease enzyme (urea amidohydrolase)</p> <p>Comment: hydrolyses urea into ammonia and carbamic acid [2].</p> <p>Comment: UreF is proposed to modulate the activation process of</p> <p>Comment: urease by eliminating the binding of nickel ions to</p> <p>Comment: noncarbamylated protein [1].</p> <p>Number of members: 20</p>   |
| Vif  |         | Retroviral Vif (Viral infectivity) protein | <p>Accession number: PF00559</p> <p>Definition: Retroviral Vif (Viral infectivity) protein</p> <p>Author: Bateman A</p> <p>Alignment method of seed: Clustalw</p> <p>Source of seed members: Swiss-Prot</p> <p>Gathering cutoffs: 25 25</p> <p>Trusted cutoffs: 53.90 53.90</p> <p>Noise cutoffs: 23.60 23.60</p> <p>HMM build command line: hmmbuild -f HMM SEED</p> <p>HMM build command line: hmmcalibrate --seed 0 HMM</p> <p>Reference Number: [1]</p> <p>Reference Medline: 95287525</p> <p>Reference Title: Aberrant Gag protein composition of a human</p> <p>Reference Title: immunodeficiency virus type 1 vif mutant produced in</p> <p>Reference Title: primary lymphocytes.</p> <p>Reference Author: Simm M, Shahabuddin M, Chao W, Allan JS, Volsky DJ;</p> <p>Reference Location: J Virol 1995;69:4582-4586.</p> <p>Database Reference: INTERPRO; IPR000475;</p> <p>Comment: -!- Human immunodeficiency virus type 1 (HIV-1) Vif is required for</p> <p>Comment: productive infection of T lymphocytes and macrophages. Virions</p> <p>Comment: produced in the absence of Vif have abnormal core morphology and</p> <p>Comment: those produced in primary T cells carry immature core proteins</p> <p>Comment: and low levels of mature capsid.</p> <p>Number of members: 503</p> |
| Vpu  |         | Vpu protein                                | <p>Accession number: PF00558</p> <p>Definition: Vpu protein</p> <p>Author: Bateman A</p> <p>Alignment method of seed: Clustalw</p> <p>Source of seed members: Swiss-Prot</p> <p>Gathering cutoffs: 15 15</p> <p>Trusted cutoffs: 15.50 15.50</p> <p>Noise cutoffs: 13.60 13.60</p> <p>HMM build command line: hmmbuild -f HMM SEED</p> <p>HMM build command line: hmmcalibrate --seed 0 HMM</p> <p>Reference Number: [1]</p> <p>Reference Medline: 97479365</p> <p>Reference Title: Enhancement of retroviral production from packaging cell</p> <p>Reference Title: lines expressing the human</p> <p>Reference Title: immunodeficiency type 1 VPU</p> <p>Reference Title: gene.</p> <p>Reference Author: Kobinger GP, Mouland AJ, Lalonde JP, Forget J, Cohen EA;</p> <p>Reference Location: Gene Ther 1997;4:868-874.</p> <p>Reference Number: [2]</p> <p>Reference Medline: 95156576</p> <p>Reference Title: The human immunodeficiency virus type 1</p> <p>Reference Title: Vpu protein</p> <p>Reference Title: specifically binds to the cytoplasmic domain of CD4:</p> <p>Reference Title: implications for the mechanism of degradation.</p> <p>Reference Author: Bour S, Schubert U, Strebel K;</p> <p>Reference Location: J Virol 1995;69:1510-1520.</p>                                |

09679203 100400

| Pfam  | Prosite   | Full Name              | Description   |
|-------|-----------|------------------------|---|
|       |           |                        | <p>Reference Number: [3]<br/> Reference Medline: 97325981<br/> Reference Title: Secondary structure and tertiary fold of the human immunodeficiency virus protein U (Vpu) cytoplasmic domain<br/> Reference Title: in solution.<br/> Reference Author: Willbold D, Hoffmann S, Rosch P;<br/> Reference Location: Eur J Biochem 1997;245:581-588.<br/> Database Reference: SCOP; 1vpu; fa; [SCOP-USA][CATH-PDBSUM]<br/> Database Reference: INTERPRO; IPR002094;<br/> Database Reference: PDB; 1vpu ; 38; 81;<br/> Database reference: PFAMB; PB003303;<br/> Database reference: PFAMB; PB005882;<br/> Comment: -!- The Vpu protein contains an N-terminal transmembrane spanning region<br/> Comment: and a C-terminal cytoplasmic region.<br/> Comment: -!- The HIV-1 Vpu protein stimulates virus production by enhancing<br/> Comment: the release of viral particles from infected cells.<br/> Comment: -!- The VPU protein binds specifically to CD4.<br/> Number of members: 194</p>  |
| XPG_N | PDOC00658 | XPG protein signatures | <p>Xeroderma pigmentosum (XP) [1] is a human autosomal recessive disease, characterized by a high incidence of sunlight-induced skin cancer. People's skin cells with this condition are hypersensitive to ultraviolet light, due to defects in the incision step of DNA excision repair. There are a minimum of seven genetic complementation groups involved in this pathway: XP-A to XP-G. The defect in XP-G can be corrected by a 133 Kd nuclear protein called XPG (or XPGC) [2].</p> <p>XPG belongs to a family of proteins [2,3,4,5,6] that are composed of two main subsets:</p> <ul style="list-style-type: none"> <li>- Subset 1, to which belongs XPG, RAD2 from budding yeast and rad13 from fission yeast. RAD2 and XPG are single-stranded DNA endonucleases [7,8]. XPG makes the 3' incision in human DNA nucleotide excision repair [9].</li> <li>- Subset 2, to which belongs mouse and human FEN-1, rad2 from fission yeast, and RAD27 from budding yeast. FEN-1 is a structure-specific endonuclease.</li> </ul> <p>In addition to the proteins listed in the above groups, this family also includes:</p> <ul style="list-style-type: none"> <li>- Fission yeast exo1, a 5'-&gt;3' double-stranded DNA exonuclease that could act in a pathway that corrects mismatched base pairs.</li> <li>- Yeast EXO1 (DHS1), a protein with probably the same function as exo1.</li> <li>- Yeast DIN7.</li> </ul> <p>Sequence alignment of this family of proteins reveals that similarities are largely confined to two regions. The first is located at the N-terminal extremity (N-region) and corresponds to the first 95 to 105 amino acids. The second region is internal (I-region) and found towards the C-terminus; it</p> |

004007 02262960



| Pfam | Prosite | Full Name | Description  |
|------|---------|-----------|--|
|      |         |           | <p>spans about 140 residues and contains a highly conserved core of 27 amino acids that includes a conserved pentapeptide (E-A-[DE]-A-[QS]). It is possible that the conserved acidic residues are involved in the catalytic mechanism of DNA excision repair in XPG. The amino acids linking the N- and I-regions are not conserved; indeed, they are largely absent from proteins belonging to the second subset.</p> <p>We have developed two signature patterns for these proteins. The first corresponds to the central part of the N-region, the second to part of the I-region and includes the putative catalytic core pentapeptide.</p> <p>Description of pattern(s) and/or profile(s)</p> <p>Consensus pattern [VI]-[KRE]-P-x-[FYIL]-V-F-D-G-x(2)-[PIL]-x-[LVC]-K<br/>Sequences known to belong to this class detected by the pattern ALL.<br/>Other sequence(s) detected in SWISS-PROT NONE.</p> <p>Consensus pattern [GS]-[LIVM]-[PER]-[FYS]-[LIVM]-x-A-P-x-E-A-[DE]-[PAS]-[QS]-[CLM]<br/>Sequences known to belong to this class detected by the pattern ALL.<br/>Other sequence(s) detected in SWISS-PROT NONE.<br/>Expert(s) to contact by email<br/>Clarkson S.G. clarkson@medecine.unige.ch</p> <p>Last update<br/>November 1997 / Patterns and text revised.</p> <p>References<br/>[ 1]<br/>Tanaka K., Wood R.D.<br/>Trends Biochem. Sci. 19:83-86(1994).</p> <p>[ 2]<br/>Scherly D., Nospikel T., Corlet J., Ucla C., Bairoch A., Clarkson S.G.<br/>Nature 363:182-185(1993).</p> <p>[ 3]<br/>Carr A.M., Sheldrick K.S., Murray J.M., Al-Harithy R., Watts F.Z., Lehmann A.R.<br/>Nucleic Acids Res. 21:1345-1349(1993).</p> <p>[ 4]<br/>Murray J.M., Tavassoli M., Al-Harithy R., Sheldrick K.S., Lehmann A.R., Carr A.M., Watts F.Z.<br/>Mol. Cell. Biol. 14:4878-4888(1994).</p> <p>[ 5]<br/>Harrington J.J., Lieber M.R.<br/>Genes Dev. 8:1344-1355(1994).</p> <p>[ 6]<br/>Szankasi P., Smith G.R.<br/>Science 267:1166-1169(1995).</p> <p>[ 7]<br/>Habraken Y., Sung P., Prakash L., Prakash S.<br/>Nature 366:365-368(1993).</p> <p>[ 8]<br/>O'Donovan A., Scherly D., Clarkson S.G., Wood R.D.<br/>J. Biol. Chem. 269:15965-15968(1994).</p> |

004007 0264960

| Pfam          | Prosite   | Full Name   | Description   |
|---------------|-----------|---|---|
|               |           |   | [ 9]<br>O'Donovan A., Davies A.A., Moggs J.G., West S.C., Wood R.D.<br>Nature 371:432-435(1994).  |
| Y_phosphatase | PDOC00323 | Tyrosine specific protein phosphatases signature and profiles | <p>Tyrosine specific protein phosphatases (EC 3.1.3.48) (PTPase) [1 to 5] are enzymes that catalyze the removal of a phosphate group attached to a tyrosine residue. These enzymes are very important in the control of cell growth, proliferation, differentiation and transformation. Multiple forms of PTPase have been characterized and can be classified into two categories: soluble PTPases and transmembrane receptor proteins that contain PTPase domain(s). The currently known PTPases are listed below:</p> <p>Soluble PTPases.</p> <ul style="list-style-type: none"> <li>- PTPN1 (PTP-1B).</li> <li>- PTPN2 (T-cell PTPase; TC-PTP).</li> <li>- PTPN3 (H1) and PTPN4 (MEG), enzymes that contain an N-terminal band 4.1-like domain (see &lt;PDOC00566&gt;) and could act at junctions between the membrane and cytoskeleton.</li> <li>- PTPN5 (STEP).</li> <li>- PTPN6 (PTP-1C; HCP; SHP) and PTPN11 (PTP-2C; SH-PTP3; Syp), enzymes which contain two copies of the SH2 domain at its N-terminal extremity. The <i>Drosophila</i> protein corkscrew (gene csw) also belongs to this subgroup.</li> <li>- PTPN7 (LC-PTP; Hematopoietic protein-tyrosine phosphatase; HePTP).</li> <li>- PTPN8 (70Z-PEP).</li> <li>- PTPN9 (MEG2).</li> <li>- PTPN12 (PTP-G1; PTP-P19).</li> </ul> <p>- Yeast PTP1.</p> <p>- Yeast PTP2 which may be involved in the ubiquitin-mediated protein degradation pathway.</p> <p>- Fission yeast pyp1 and pyp2 which play a role in inhibiting the onset of mitosis.</p> <p>- Fission yeast pyp3 which contributes to the dephosphorylation of cdc2.</p> <p>- Yeast CDC14 which may be involved in chromosome segregation.</p> <p>- <i>Yersinia</i> virulence plasmid PTPases (gene yopH).</p> <p>- <i>Autographa californica</i> nuclear polyhedrosis virus 19 Kd PTPase.</p> <p>Dual specificity PTPases.</p> <ul style="list-style-type: none"> <li>- DUSP1 (PTPN10; MAP kinase phosphatase-1; MKP-1); which dephosphorylates MAP kinase on both Thr-183 and Tyr-185.</li> <li>- DUSP2 (PAC-1), a nuclear enzyme that dephosphorylates MAP kinases ERK1 and ERK2 on both Thr and Tyr residues.</li> <li>- DUSP3 (VHR).</li> <li>- DUSP4 (HvH2).</li> <li>- DUSP5 (HvH3).</li> <li>- DUSP6 (Pyst1; MKP-3).</li> <li>- DUSP7 (Pyst2; MKP-X).</li> <li>- Yeast MSG5, a PTPase that dephosphorylates MAP kinase FUS3.</li> <li>- Yeast YVH1.</li> <li>- <i>Vaccinia</i> virus H1 PTPase; a dual specificity phosphatase.</li> </ul> <p>Receptor PTPases.</p> |

004007 00264960

| Pfam                                  | Prosite       | Full Name     | Description  |        |               |               |  |  |  |  |  |         |     |     |        |  |  |                                       |   |   |   |   |   |  |                                 |   |   |   |   |   |  |                 |   |   |   |   |   |  |                 |   |   |   |   |   |  |                 |   |   |   |   |   |  |          |   |    |   |   |   |  |           |   |   |   |   |   |  |           |   |    |   |   |   |  |             |   |   |   |   |   |  |           |   |   |   |   |   |  |        |   |   |   |   |   |  |          |   |   |   |   |   |  |
|---------------------------------------|---------------|---------------|--|--------|---------------|---------------|--|--|--|--|--|---------|-----|-----|--------|--|--|---------------------------------------|---|---|---|---|---|--|---------------------------------|---|---|---|---|---|--|-----------------|---|---|---|---|---|--|-----------------|---|---|---|---|---|--|-----------------|---|---|---|---|---|--|----------|---|----|---|---|---|--|-----------|---|---|---|---|---|--|-----------|---|----|---|---|---|--|-------------|---|---|---|---|---|--|-----------|---|---|---|---|---|--|--------|---|---|---|---|---|--|----------|---|---|---|---|---|--|
|                                       |               |               | <p>Structurally, all known receptor PTPases, are made up of a variable length extracellular domain, followed by a transmembrane region and a C-terminal catalytic cytoplasmic domain. Some of the receptor PTPases contain fibronectin type III (FN-III) repeats, immunoglobulin-like domains, MAM domains or carbonic anhydrase-like domains in their extracellular region. The cytoplasmic region generally contains two copies of the PTPase domain. The first seems to have enzymatic activity, while the second is inactive but seems to affect substrate specificity of the first. In these domains, the catalytic cysteine is generally conserved but some other, presumably important, residues are not.</p> <p>In the following table, the domain structure of known receptor PTPases is shown:</p> <table><tr><th></th><th>Extracellular</th><th colspan="5">Intracellular</th></tr><tr><th></th><th>Ig FN-3</th><th>CAH</th><th>MAM</th><th colspan="3">PTPase</th></tr><tr><td>Leukocyte common antigen (LCA) (CD45)</td><td>0</td><td>2</td><td>0</td><td>0</td><td>2</td><td></td></tr><tr><td>Leukocyte antigen related (LAR)</td><td>3</td><td>8</td><td>0</td><td>0</td><td>2</td><td></td></tr><tr><td>Drosophila DLAR</td><td>3</td><td>9</td><td>0</td><td>0</td><td>2</td><td></td></tr><tr><td>Drosophila DPTP</td><td>2</td><td>2</td><td>0</td><td>0</td><td>2</td><td></td></tr><tr><td>PTP-alpha (LRP)</td><td>0</td><td>0</td><td>0</td><td>0</td><td>2</td><td></td></tr><tr><td>PTP-beta</td><td>0</td><td>16</td><td>0</td><td>0</td><td>1</td><td></td></tr><tr><td>PTP-gamma</td><td>0</td><td>1</td><td>1</td><td>0</td><td>2</td><td></td></tr><tr><td>PTP-delta</td><td>0</td><td>&gt;7</td><td>0</td><td>0</td><td>2</td><td></td></tr><tr><td>PTP-epsilon</td><td>0</td><td>0</td><td>0</td><td>0</td><td>2</td><td></td></tr><tr><td>PTP-kappa</td><td>1</td><td>4</td><td>0</td><td>1</td><td>2</td><td></td></tr><tr><td>PTP-mu</td><td>1</td><td>4</td><td>0</td><td>1</td><td>2</td><td></td></tr><tr><td>PTP-zeta</td><td>0</td><td>1</td><td>1</td><td>0</td><td>2</td><td></td></tr></table> <p>PTPase domains consist of about 300 amino acids. There are two conserved cysteines, the second one has been shown to be absolutely required for activity. Furthermore, a number of conserved residues in its immediate vicinity have also been shown to be important.</p> <p>We derived a signature pattern for PTPase domains centered on the active site cysteine.</p> <p>There are three profiles for PTPases, the first one spans the complete domain and is not specific to any subtype. The second profile is specific to dual-specificity PTPases and the third one to the PTP subfamily.</p> <p>Description of pattern(s) and/or profile(s)</p> <p>Consensus pattern [LIVMF]-H-C-x(2)-G-x(3)-[STC]-[STAGP]-x-[LIVMFY] [C is the active site residue]<br/>Sequences known to belong to this class detected by the pattern ALL, except for nine sequences.<br/>Other sequence(s) detected in SWISS-PROT 3.</p> <p>Sequences known to belong to this class detected by the 1st profile ALL.<br/>Other sequence(s) detected in SWISS-PROT 2.</p> <p>Sequences known to belong to this class detected by the 2nd</p> |        | Extracellular | Intracellular |  |  |  |  |  | Ig FN-3 | CAH | MAM | PTPase |  |  | Leukocyte common antigen (LCA) (CD45) | 0 | 2 | 0 | 0 | 2 |  | Leukocyte antigen related (LAR) | 3 | 8 | 0 | 0 | 2 |  | Drosophila DLAR | 3 | 9 | 0 | 0 | 2 |  | Drosophila DPTP | 2 | 2 | 0 | 0 | 2 |  | PTP-alpha (LRP) | 0 | 0 | 0 | 0 | 2 |  | PTP-beta | 0 | 16 | 0 | 0 | 1 |  | PTP-gamma | 0 | 1 | 1 | 0 | 2 |  | PTP-delta | 0 | >7 | 0 | 0 | 2 |  | PTP-epsilon | 0 | 0 | 0 | 0 | 2 |  | PTP-kappa | 1 | 4 | 0 | 1 | 2 |  | PTP-mu | 1 | 4 | 0 | 1 | 2 |  | PTP-zeta | 0 | 1 | 1 | 0 | 2 |  |
|                                       | Extracellular | Intracellular |  |        |               |               |  |  |  |  |  |         |     |     |        |  |  |                                       |   |   |   |   |   |  |                                 |   |   |   |   |   |  |                 |   |   |   |   |   |  |                 |   |   |   |   |   |  |                 |   |   |   |   |   |  |          |   |    |   |   |   |  |           |   |   |   |   |   |  |           |   |    |   |   |   |  |             |   |   |   |   |   |  |           |   |   |   |   |   |  |        |   |   |   |   |   |  |          |   |   |   |   |   |  |
|                                       | Ig FN-3       | CAH           | MAM  | PTPase |               |               |  |  |  |  |  |         |     |     |        |  |  |                                       |   |   |   |   |   |  |                                 |   |   |   |   |   |  |                 |   |   |   |   |   |  |                 |   |   |   |   |   |  |                 |   |   |   |   |   |  |          |   |    |   |   |   |  |           |   |   |   |   |   |  |           |   |    |   |   |   |  |             |   |   |   |   |   |  |           |   |   |   |   |   |  |        |   |   |   |   |   |  |          |   |   |   |   |   |  |
| Leukocyte common antigen (LCA) (CD45) | 0             | 2             | 0  | 0      | 2             |               |  |  |  |  |  |         |     |     |        |  |  |                                       |   |   |   |   |   |  |                                 |   |   |   |   |   |  |                 |   |   |   |   |   |  |                 |   |   |   |   |   |  |                 |   |   |   |   |   |  |          |   |    |   |   |   |  |           |   |   |   |   |   |  |           |   |    |   |   |   |  |             |   |   |   |   |   |  |           |   |   |   |   |   |  |        |   |   |   |   |   |  |          |   |   |   |   |   |  |
| Leukocyte antigen related (LAR)       | 3             | 8             | 0  | 0      | 2             |               |  |  |  |  |  |         |     |     |        |  |  |                                       |   |   |   |   |   |  |                                 |   |   |   |   |   |  |                 |   |   |   |   |   |  |                 |   |   |   |   |   |  |                 |   |   |   |   |   |  |          |   |    |   |   |   |  |           |   |   |   |   |   |  |           |   |    |   |   |   |  |             |   |   |   |   |   |  |           |   |   |   |   |   |  |        |   |   |   |   |   |  |          |   |   |   |   |   |  |
| Drosophila DLAR                       | 3             | 9             | 0  | 0      | 2             |               |  |  |  |  |  |         |     |     |        |  |  |                                       |   |   |   |   |   |  |                                 |   |   |   |   |   |  |                 |   |   |   |   |   |  |                 |   |   |   |   |   |  |                 |   |   |   |   |   |  |          |   |    |   |   |   |  |           |   |   |   |   |   |  |           |   |    |   |   |   |  |             |   |   |   |   |   |  |           |   |   |   |   |   |  |        |   |   |   |   |   |  |          |   |   |   |   |   |  |
| Drosophila DPTP                       | 2             | 2             | 0  | 0      | 2             |               |  |  |  |  |  |         |     |     |        |  |  |                                       |   |   |   |   |   |  |                                 |   |   |   |   |   |  |                 |   |   |   |   |   |  |                 |   |   |   |   |   |  |                 |   |   |   |   |   |  |          |   |    |   |   |   |  |           |   |   |   |   |   |  |           |   |    |   |   |   |  |             |   |   |   |   |   |  |           |   |   |   |   |   |  |        |   |   |   |   |   |  |          |   |   |   |   |   |  |
| PTP-alpha (LRP)                       | 0             | 0             | 0  | 0      | 2             |               |  |  |  |  |  |         |     |     |        |  |  |                                       |   |   |   |   |   |  |                                 |   |   |   |   |   |  |                 |   |   |   |   |   |  |                 |   |   |   |   |   |  |                 |   |   |   |   |   |  |          |   |    |   |   |   |  |           |   |   |   |   |   |  |           |   |    |   |   |   |  |             |   |   |   |   |   |  |           |   |   |   |   |   |  |        |   |   |   |   |   |  |          |   |   |   |   |   |  |
| PTP-beta                              | 0             | 16            | 0  | 0      | 1             |               |  |  |  |  |  |         |     |     |        |  |  |                                       |   |   |   |   |   |  |                                 |   |   |   |   |   |  |                 |   |   |   |   |   |  |                 |   |   |   |   |   |  |                 |   |   |   |   |   |  |          |   |    |   |   |   |  |           |   |   |   |   |   |  |           |   |    |   |   |   |  |             |   |   |   |   |   |  |           |   |   |   |   |   |  |        |   |   |   |   |   |  |          |   |   |   |   |   |  |
| PTP-gamma                             | 0             | 1             | 1  | 0      | 2             |               |  |  |  |  |  |         |     |     |        |  |  |                                       |   |   |   |   |   |  |                                 |   |   |   |   |   |  |                 |   |   |   |   |   |  |                 |   |   |   |   |   |  |                 |   |   |   |   |   |  |          |   |    |   |   |   |  |           |   |   |   |   |   |  |           |   |    |   |   |   |  |             |   |   |   |   |   |  |           |   |   |   |   |   |  |        |   |   |   |   |   |  |          |   |   |   |   |   |  |
| PTP-delta                             | 0             | >7            | 0  | 0      | 2             |               |  |  |  |  |  |         |     |     |        |  |  |                                       |   |   |   |   |   |  |                                 |   |   |   |   |   |  |                 |   |   |   |   |   |  |                 |   |   |   |   |   |  |                 |   |   |   |   |   |  |          |   |    |   |   |   |  |           |   |   |   |   |   |  |           |   |    |   |   |   |  |             |   |   |   |   |   |  |           |   |   |   |   |   |  |        |   |   |   |   |   |  |          |   |   |   |   |   |  |
| PTP-epsilon                           | 0             | 0             | 0  | 0      | 2             |               |  |  |  |  |  |         |     |     |        |  |  |                                       |   |   |   |   |   |  |                                 |   |   |   |   |   |  |                 |   |   |   |   |   |  |                 |   |   |   |   |   |  |                 |   |   |   |   |   |  |          |   |    |   |   |   |  |           |   |   |   |   |   |  |           |   |    |   |   |   |  |             |   |   |   |   |   |  |           |   |   |   |   |   |  |        |   |   |   |   |   |  |          |   |   |   |   |   |  |
| PTP-kappa                             | 1             | 4             | 0  | 1      | 2             |               |  |  |  |  |  |         |     |     |        |  |  |                                       |   |   |   |   |   |  |                                 |   |   |   |   |   |  |                 |   |   |   |   |   |  |                 |   |   |   |   |   |  |                 |   |   |   |   |   |  |          |   |    |   |   |   |  |           |   |   |   |   |   |  |           |   |    |   |   |   |  |             |   |   |   |   |   |  |           |   |   |   |   |   |  |        |   |   |   |   |   |  |          |   |   |   |   |   |  |
| PTP-mu                                | 1             | 4             | 0  | 1      | 2             |               |  |  |  |  |  |         |     |     |        |  |  |                                       |   |   |   |   |   |  |                                 |   |   |   |   |   |  |                 |   |   |   |   |   |  |                 |   |   |   |   |   |  |                 |   |   |   |   |   |  |          |   |    |   |   |   |  |           |   |   |   |   |   |  |           |   |    |   |   |   |  |             |   |   |   |   |   |  |           |   |   |   |   |   |  |        |   |   |   |   |   |  |          |   |   |   |   |   |  |
| PTP-zeta                              | 0             | 1             | 1  | 0      | 2             |               |  |  |  |  |  |         |     |     |        |  |  |                                       |   |   |   |   |   |  |                                 |   |   |   |   |   |  |                 |   |   |   |   |   |  |                 |   |   |   |   |   |  |                 |   |   |   |   |   |  |          |   |    |   |   |   |  |           |   |   |   |   |   |  |           |   |    |   |   |   |  |             |   |   |   |   |   |  |           |   |   |   |   |   |  |        |   |   |   |   |   |  |          |   |   |   |   |   |  |

004001" E0264960

| Pfam   | Prosite | Full Name                 | Description   |
|--------|---------|---------------------------|---|
|        |         |                           | <p>profile ALL dual type PTPases.<br/>Other sequence(s) detected in SWISS-PROT NONE.</p> <p>Sequences known to belong to this class detected by the 3rd profile ALL PTP type PTPases.<br/>Other sequence(s) detected in SWISS-PROT NONE.</p> <p>Note the M-phase inducer phosphatases (cdc25-type phosphatase) are tyrosine- protein phosphatases that are not structurally related to the above PTPases.</p> <p>Note this documentation entry is linked to both a signature pattern and to profiles. As profiles are much more sensitive than the pattern, you should use them if you have access to the necessary software tools to do so.</p> <p>Last update<br/>July 1999 / Text revised.</p> <p>References<br/>[ 1]<br/>Fischer E.H., Charbonneau H., Tonks N.K.<br/>Science 253:401-406(1991).</p> <p>[ 2]<br/>Charbonneau H., Tonks N.K.<br/>Annu. Rev. Cell Biol. 8:463-493(1992).</p> <p>[ 3]<br/>Trowbridge I.S.<br/>J. Biol. Chem. 266:23517-23520(1991).</p> <p>[ 4]<br/>Tonks N.K., Charbonneau H.<br/>Trends Biochem. Sci. 14:497-500(1989).</p> <p>[ 5]<br/>Hunter T.<br/>Cell 58:1013-1016(1989).</p> |
| Zein   |         | Zein seed storage protein | <p>Accession number: PF01559<br/>Definition: Zein seed storage protein<br/>Author: Bateman A<br/>Alignment method of seed: Clustalw<br/>Source of seed members: Pfam-B_181 (release 4.0)<br/>Gathering cutoffs: -21 -21<br/>Trusted cutoffs: 4.60 4.60<br/>Noise cutoffs: -46.60 -46.60<br/>HMM build command line: hmmbuild -F HMM SEED<br/>HMM build command line: hmmscalibrate --seed 0 HMM<br/>Reference Number: [1]<br/>Reference Medline: 93197294<br/>Reference Title: Studies of the zein-like alpha-prolamins based on an<br/>Reference Title: analysis of amino acid sequences: implications for their<br/>Reference Title: evolution and three-dimensional structure.<br/>Reference Author: Garratt R, Oliva G, Caracelli I, Leite A, Arruda P;<br/>Reference Location: Proteins 1993;15:88-99.<br/>Database Reference: INTERPRO; IPR002530;<br/>Comment: Zeins are seed storage proteins. They are unusually rich in<br/>Comment: glutamine, proline, alanine, and leucine residues and their<br/>Comment: sequences show a series of tandem repeats [1].<br/>Number of members: 48</p>                  |
| zf-AN1 |         | AN1-like Zinc finger      | <p>Accession number: PF01428<br/>Definition: AN1-like Zinc finger<br/>Author: Bateman A, SMART<br/>Alignment method of seed: Manual<br/>Source of seed members: SMART<br/>Gathering cutoffs: 16 16<br/>Trusted cutoffs: 16.40 16.40</p>   |

004007 00264960

| Pfam        | Prosite   | Full Name                   | Description   |
|-------------|-----------|-----------------------------|---|
|             |           |                             | <p>Noise cutoffs: 7.30 7.30</p> <p>HMM build command line: hmmbuild HMM SEED</p> <p>HMM build command line: hmmscalibrate --seed 0 HMM</p> <p>Reference Number: [1]</p> <p>Reference Medline: 93292985</p> <p>Reference Title: Two related localized mRNAs from <i>Xenopus laevis</i> encode</p> <p>Reference Title: ubiquitin-like fusion proteins.</p> <p>Reference Author: Linnen JM, Bailey CP, Weeks DL;</p> <p>Reference Location: Gene 1993;128:181-188.</p> <p>Database reference: SMART; Znf_AN1;</p> <p>Database Reference INTERPRO; IPR000058;</p> <p>Comment: Zinc finger at the C-terminus of An1</p> <p>Swiss:Q91889, a ubiquitin-like</p> <p>Comment: protein in <i>Xenopus laevis</i>.</p> <p>Comment: The following pattern describes the zinc finger.</p> <p>Comment: C-X2-C-X(9-12)-C-X(1-2)-C-X4-C-X2-H-X5-H-X-C</p> <p>Comment: Where X can be any amino acid, and numbers in brackets</p> <p>Comment: indicate the number of residues.</p> <p>Number of members: 18</p> |
| zf-B_box    | PDOC50015 | B-box zinc finger           | <p>Accession number: PF00643</p> <p>Definition: B-box zinc finger.</p> <p>Author: Bateman A</p> <p>Alignment method of seed: pftools</p> <p>Source of seed members: Prosite</p> <p>Gathering cutoffs: 25 25</p> <p>Trusted cutoffs: 26.00 26.00</p> <p>Noise cutoffs: 24.50 29.90</p> <p>HMM build command line: hmmbuild HMM SEED</p> <p>HMM build command line: hmmscalibrate --seed 0 HMM</p> <p>Database Reference: SCOP; 1fre; fa; [SCOP-USA][CATH-PDBSUM]</p> <p>Database reference: PROSITE_PROFILE; PS50119;</p> <p>Database Reference: PROSITE; PDOC50015</p> <p>Database Reference INTERPRO; IPR002991;</p> <p>Database Reference PDB; 1fre ; 4; 42;</p> <p>Database reference: PFAMB; PB002777;</p> <p>Database reference: PFAMB; PB010625;</p> <p>Database reference: PFAMB; PB041771;</p> <p>Number of members: 44</p>   |
| zf-CONSTANS |           | CONSTANS family zinc finger | <p>Accession number: PF01760</p> <p>Definition: CONSTANS family zinc finger</p> <p>Author: Bateman A</p> <p>Alignment method of seed: Clustalw</p> <p>Source of seed members: Pfam-B_1072 (release 4.2)</p> <p>Gathering cutoffs: 25 10</p> <p>Trusted cutoffs: 76.10 17.20</p> <p>Noise cutoffs: 9.70 9.70</p> <p>HMM build command line: hmmbuild HMM SEED</p> <p>HMM build command line: hmmscalibrate --seed 0 HMM</p> <p>Reference Number: [1]</p> <p>Reference Medline: 95211836</p> <p>Reference Title: The CONSTANS gene of <i>Arabidopsis</i> promotes flowering and</p> <p>Reference Title: encodes a protein showing similarities to zinc finger</p> <p>Reference Title: transcription factors.</p> <p>Reference Author: Putterill J, Robson F, Lee K, Simon R, Coupland G;</p> <p>Reference Location: Cell 1995;80:847-857.</p> <p>Database Reference INTERPRO; IPR002926;</p> <p>Number of members: 45</p>   |
| zf-DHHC     |           | DHHC zinc finger domain     | <p>Accession number: PF01529</p> <p>Definition: DHHC zinc finger domain</p> <p>Author: Bateman A</p> <p>Alignment method of seed: Clustalw</p> <p>Source of seed members: Pfam-B_945 (release 4.0)</p> <p>Gathering cutoffs: 22 22</p>  |

004001 00264960

| Pfam    | Prosite | Full Name   | Description   |
|---------|---------|-------------|---|
|         |         |             | <p>Trusted cutoffs: 22.40 22.40<br/> Noise cutoffs: -22.40 -22.40<br/> HMM build command line: hmmbuild HMM SEED<br/> HMM build command line: hmmscalibrate --seed 0 HMM<br/> Reference Number: [1]<br/> Reference Medline: 99250263<br/> Reference Title: The drosophila STAM gene homolog is in a tight gene<br/> Reference Title: cluster, and its expression correlates to that of the<br/> Reference Title: adjacent gene ial.<br/> Reference Author: Mesilaty-Gross S, Reich A, Motro B, Wides R;<br/> Reference Location: Gene 1999;231:173-186.<br/> Reference Number: [2]<br/> Reference Medline: 97315340<br/> Reference Title: Variations of the C2H2 zinc finger motif in the yeast<br/> Reference Title: genome and classification of yeast zinc finger proteins.<br/> Reference Author: Bohm S, Frishman D, Mewes HW;<br/> Reference Location: Nucleic Acids Res 1997;25:2464-2469.<br/> Reference Number: [3]<br/> Reference Medline: 99321009<br/> Reference Title: The DHHC domain: a new highly conserved cysteine-rich<br/> Reference Title: motif.<br/> Reference Author: Putilina T, Wong P, Gentleman S;<br/> Reference Location: Mol Cell Biochem 1999;195:219-226.<br/> Reference Number: [4]<br/> Reference Medline: 10490616<br/> Reference Title: Erf2, a Novel Gene Product That Affects the Localization<br/> Reference Title: and Palmitoylation of Ras2 in Saccharomyces cerevisiae.<br/> Reference Author: Bartels DJ, Mitchell DA, Dong X, Deschenes RJ;<br/> Reference Location: Mol Cell Biol 1999;19:6775-6787.<br/> Database Reference: INTERPRO; IPR001594;<br/> Comment: This domain is also known as NEW1 [2].<br/> This domain is<br/> Comment: predicted to be a zinc binding domain. The function<br/> Comment: of this domain is unknown, but it has been predicted to<br/> Comment: be involved in protein-protein or protein-DNA interactions [3].<br/> Number of members: 34</p> |
| zf-MYND |         | MYND finger | <p>Accession number: PF01753<br/> Definition: MYND finger<br/> Author: Bateman A<br/> Alignment method of seed: Manual<br/> Source of seed members: Bateman A<br/> Gathering cutoffs: 11 11<br/> Trusted cutoffs: 17.30 17.30<br/> Noise cutoffs: 5.50 5.50<br/> HMM build command line: hmmbuild HMM SEED<br/> HMM build command line: hmmscalibrate --seed 0 HMM<br/> Reference Number: [1]<br/> Reference Medline: 96203118<br/> Reference Title: DEAF-1, a novel protein that binds an essential region in a<br/> Reference Title: Deformed response element.<br/> Reference Author: Gross CT, McGinnis W;<br/> Reference Location: EMBO J 1996;15:1961-1970.<br/> Reference Number: [2]<br/> Reference Medline: 98079069<br/> Reference Title: Molecular cloning, sequence analysis, expression, and<br/> Reference Title: tissue distribution of suppressin, a novel suppressor of<br/> Reference Title: cell cycle entry.<br/> Reference Author: LeBoeuf RD, Ban EM, Green MM, Stone</p>   |

| Pfam         | Prosite   | Full Name   | Description   |
|--------------|-----------|---|---|
|              |           |   | <p>AS, Propst SM, Blalock</p> <p>Reference Author: JE, Tauber JD;</p> <p>Reference Location: J Biol Chem 1998;273:361-368.</p> <p>Database Reference INTERPRO; IPR002893;</p> <p>Number of members: 48</p>  |
| Zn_carbOpept | PDOC00123 | Zinc carboxypeptidases, zinc-binding regions signatures | <p>There are a number of different types of zinc-dependent carboxypeptidases (EC 3.4.17.-) [1,2]. All these enzymes seem to be structurally and functionally related. The enzymes that belong to this family are listed below.</p> <ul style="list-style-type: none"> <li>- Carboxypeptidase A1 (EC 3.4.17.1), a pancreatic digestive enzyme that can remove all C-terminal amino acids with the exception of Arg, Lys and Pro.</li> <li>- Carboxypeptidase A2 (EC 3.4.17.15), a pancreatic digestive enzyme with a specificity similar to that of carboxypeptidase A1, but with a preference for bulkier C-terminal residues.</li> <li>- Carboxypeptidase B (EC 3.4.17.2), also a pancreatic digestive enzyme, but that preferentially removes C-terminal Arg and Lys.</li> <li>- Carboxypeptidase N (EC 3.4.17.3) (also known as arginine carboxypeptidase), a plasma enzyme which protects the body from potent vasoactive and inflammatory peptides containing C-terminal Arg or Lys (such as kinins or anaphylatoxins) which are released into the circulation.</li> <li>- Carboxypeptidase H (EC 3.4.17.10) (also known as enkephalin convertase or carboxypeptidase E), an enzyme located in secretory granules of pancreatic islets, adrenal gland, pituitary and brain. This enzyme removes residual C-terminal Arg or Lys remaining after initial endoprotease cleavage during prohormone processing.</li> <li>- Carboxypeptidase M (EC 3.4.17.12), a membrane bound Arg and Lys specific enzyme.</li> </ul> <p>It is ideally situated to act on peptide hormones at local tissue sites where it could control their activity before or after interaction with specific plasma membrane receptors.</p> <ul style="list-style-type: none"> <li>- Mast cell carboxypeptidase (EC 3.4.17.1), an enzyme with a specificity to carboxypeptidase A, but found in the secretory granules of mast cells.</li> <li>- Streptomyces griseus carboxypeptidase (Cpase SG) (EC 3.4.17.-) [3], which combines the specificities of mammalian carboxypeptidases A and B.</li> <li>- Thermoactinomyces vulgaris carboxypeptidase T (EC 3.4.17.18) (CPT) [4], which also combines the specificities of carboxypeptidases A and B.</li> <li>- AEBP1 [5], a transcriptional repressor active in preadipocytes. AEBP1 seems to regulate transcription by cleavage of other transcriptional proteins.</li> <li>- Yeast hypothetical protein YHR132c.</li> </ul> <p>All of these enzymes bind an atom of zinc. Three conserved residues are implicated in the binding of the zinc atom: two histidines and a glutamic acid. We have derived two signature patterns which contain these three zinc-ligands.</p> |

004001 00264960

| Pfam | Prosite | Full Name                                   | Description   |
|------|---------|---|---|
|      |         |   | <p>Description of pattern(s) and/or profile(s)</p> <p>Consensus pattern [PK]-x-[LIVMFY]-x-[LIVMFY]-x(4)-H-[STAG]-x-E-x-[LIVM]-[STAG]-x(6)-[LIVMFYTA] [H and E are zinc ligands]<br/>Sequences known to belong to this class detected by the pattern ALL.<br/>Other sequence(s) detected in SWISS-PROT Bacillus sphaericus endopeptidase I which hydrolyses the gamma-D-Glu-(L)meso-diaminopimelic acid bond of spore cortex peptidoglycan [6] and which is possibly distantly related to zinc carboxypeptidases.</p> <p>Consensus pattern H-[STAG]-x(3)-[LIVME]-x(2)-[LIVMFYW]-P-[FYW] [H is a zinc ligand]<br/>Sequences known to belong to this class detected by the pattern ALL.<br/>Other sequence(s) detected in SWISS-PROT 40.</p> <p>Note if a protein includes both signatures, the probability of it being a eukaryotic zinc carboxypeptidase is 100%</p> <p>Note these proteins belong to families M14A/M14B in the classification of peptidases [7,E1].<br/>Last update<br/>November 1995 / Patterns and text revised.<br/>References<br/>[ 1]<br/>Tan F., Chan S.J., Steiner D.F., Schilling J.W., Skidgel R.A.<br/>J. Biol. Chem. 264:13165-13170(1989).<br/>[ 2]<br/>Reynolds D.S., Stevens R.L., Gurley D.S., Lane W.S., Austen K.F., Serafin W.E.<br/>J. Biol. Chem. 264:20094-20099(1989).<br/>[ 3]<br/>Narahashi Y.<br/>J. Biochem. 107:879-886(1990).<br/>[ 4]<br/>Teplyakov A., Polyakov K., Obmolova G., Strokopytov B., Kuranova I., Osterman A.L., Grishin N.V., Smulevitch S.V., Zagnitko O.P., Galperina O.V., Matz M.V., Stepanov V.M.<br/>Eur. J. Biochem. 208:281-288(1992).<br/>[ 5]<br/>He G.-P., Muise A., Li A.W., Ro H.-S.<br/>Nature 378:92-96(1995).<br/>[ 6]<br/>Hourdou M.-L., Guinand M., Vacheron M.J., Michel G., Denoroy L., Duez C.M., Englebert S., Joris B., Weber G., Ghuyssen J.-M.<br/>Biochem. J. 292:563-570(1993).<br/>[ 7]<br/>Rawlings N.D., Barrett A.J.<br/>Meth. Enzymol. 248:183-228(1995).<br/>[E1]<br/><a href="http://www.expasy.ch/cgi-bin/lists?peptidas.txt">http://www.expasy.ch/cgi-bin/lists?peptidas.txt</a></p> |
| ZZ   |         | Zinc finger present in dystrophin, CBP/p300 | <p>Accession number: PF00569<br/>Definition: Zinc finger present in dystrophin, CBP/p300<br/>Author: SMART<br/>Alignment method of seed: Manual<br/>Source of seed members: Alignment kindly provided by SMART<br/>Gathering cutoffs: 14 14<br/>Trusted cutoffs: 14.60 14.60<br/>Noise cutoffs: 10.90 10.90<br/>HMM build command line: hmmbuild HMM SEED<br/>HMM build command line: hmmbuild --seed 0 HMM<br/>Reference Number: [1]<br/>Reference Medline: 96402609<br/>Reference Title: ZZ and TAZ: new putative zinc fingers in</p>   |

004007 02264960



1041

| Pfam | Prosite | Full Name | Description  |
|------|---------|-----------|--|
|      |         |           | dystrophin and<br>Reference Title: other proteins.<br>Reference Author: Ponting CP, Blake DJ, Davies KE,<br>Kendrick-Jones J, Winder<br>Reference Author: SJ;<br>Reference Location: Trends Biochem Sci 1996;21:11-13.<br>Database Reference: EXPERT; Chris.Ponting@human-<br>anatomy.oxford.ac.uk;<br>Database Reference INTERPRO; IPR000433;<br>Database reference: PFAMB; PB041629;<br>Comment: ZZ in dystrophin binds calmodulin<br>Comment: Putative zinc finger; binding not yet shown.<br>Number of members: 87 |

004007" E026/960

AA. Activities of Polypeptides Comprising Signal Peptides

Polypeptides comprising signal peptides are a family of proteins that are typically targeted to (1) a particular organelle or intracellular compartment, (2) interact with a particular molecule or (3) for secretion outside of a host cell. Example of polypeptides comprising signal peptides include, without limitation, secreted proteins, soluble proteins, receptors, proteins retained in the ER, etc.

These proteins comprising signal peptides are useful to modulate ligand-receptor interactions, cell-to-cell communication, signal transduction, intracellular communication, and activities and/or chemical cascades that take part in an organism outside or within of any particular cell.

One class of such proteins are soluble proteins which are transported out of the cell. These proteins can act as ligands that bind to receptor to trigger signal transduction or to permit communication between cells.

Another class is receptor proteins which also comprise a retention domain that lodges the receptor protein in the membrane when the cell transports the receptor to the surface of the cell. Like the soluble ligands, receptors can also modulate signal transduction and communication between cells.

In addition the signal peptide itself can serve as a ligand for some receptors. An example is the interaction of the ER targeting signal peptide with the signal recognition particle (SRP). Here, the SRP binds to the signal peptide, halting translation, and the resulting SRP complex then binds to docking proteins located on the surface of the ER, prompting transfer of the protein into the ER.

A description of signal peptide residue composition is described below in Subsection IV.C.1.

### III. Methods of Modulating Polypeptide Production

It is contemplated that polynucleotides of the invention can be incorporated into a host cell or in-vitro system to modulate polypeptide production. For instance, the SDFs prepared as described herein can be used to prepare expression cassettes useful in a number of techniques for suppressing or enhancing expression.

An example are polynucleotides comprising sequences to be transcribed, such as coding sequences, of the present invention can be inserted into nucleic acid constructs to modulate polypeptide production. Typically, such sequences to be transcribed are heterologous to at least one element of the nucleic acid construct to generate a chimeric gene or construct.

Another example of useful polynucleotides are nucleic acid molecules comprising regulatory sequences of the present invention. Chimeric genes or constructs can be generated when the regulatory sequences of the invention linked to heterologous sequences in a vector construct. Within the scope of invention are such chimeric gene and/or constructs.

Also within the scope of the invention are nucleic acid molecules, whereof at least a part or fragment of these DNA molecules are presented in TABLE 1 of the present application, and wherein the coding sequence is under the control of its own promoter and/or its own regulatory elements. Such molecules are useful for transforming the genome of a host cell or an organism regenerated from said host cell for modulating polypeptide production.

Additionally, a vector capable of producing the oligonucleotide can be inserted into the host cell to deliver the oligonucleotide.

More detailed description of components to be included in vector constructs are described both above and below.

Whether the chimeric vectors or native nucleic acids are utilized, such polynucleotides can be incorporated into a host cell to modulate polypeptide production. Native genes and/or nucleic acid molecules can be effective when exogenous to the host cell.

Methods of modulating polypeptide expression includes, without limitation:

Suppression methods, such as

Antisense

Ribozymes

Co-suppression

Insertion of Sequences into the Gene to be Modulated

Regulatory Sequence Modulation.

as well as Methods for Enhancing Production, such as  
Insertion of Exogenous Sequences; and  
Regulatory Sequence Modulation.

5        III.A. Suppression

Expression cassettes of the invention can be used to suppress expression of  
endogenous genes which comprise the SDF sequence. Inhibiting expression can be useful,  
for instance, to tailor the ripening characteristics of a fruit (Oeller et al., *Science* 254:437  
(1991)) or to influence seed size\_(WO98/07842) or to provoke cell ablation (Mariani et al.,  
10    Nature 357: 384-387 (1992)).

As described above, a number of methods can be used to inhibit gene expression in  
plants, such as antisense, ribozyme, introduction of exogenous genes into a host cell,  
insertion of a polynucleotide sequence into the coding sequence and/or the promoter of the  
endogenous gene of interest, and the like.

15        III.A.1. Antisense

An expression cassette as described above can be transformed into host cell or  
plant to produce an antisense strand of RNA. For plant cells, antisense RNA inhibits gene  
expression by preventing the accumulation of mRNA which encodes the enzyme of interest, *see*,  
e.g., Sheehy et al., *Proc. Nat. Acad. Sci. USA*, 85:8805 (1988), and Hiatt et al., U.S. Patent No.  
20    4,801,340.

III.A.2. Ribozymes

Similarly, ribozyme constructs can be transformed into a plant to cleave mRNA  
and down-regulate translation.

III.A.3. Co-Suppression

25        Another method of suppression is by introducing an exogenous copy of the gene  
to be suppressed. Introduction of expression cassettes in which a nucleic acid is configured in  
the sense orientation with respect to the promoter has been shown to prevent the accumulation of  
mRNA. A detailed description of this method is described above.

III.A.4.        Insertion of Sequences into the Gene to be Modulated

09679203-100400  
004007-02262960

Yet another means of suppressing gene expression is to insert a polynucleotide into the gene of interest to disrupt transcription or translation of the gene.

Homologous recombination could be used to target a polynucleotide insert to a gene using the Cre-Lox system (A.C. Vergunst et al., *Nucleic Acids Res.* 26:2729 (1998), A.C. Vergunst et al., *Plant Mol. Biol.* 38:393 (1998), H. Albert et al., *Plant J.* 7:649 (1995)).

In addition, random insertion of polynucleotides into a host cell genome can also be used to disrupt the gene of interest. Azpiroz-Leehan et al., *Trends in Genetics* 13:152 (1997). In this method, screening for clones from a library containing random insertions is preferred for identifying those that have polynucleotides inserted into the gene of interest. Such screening can be performed using probes and/or primers described above based on sequences from TABLE 1, fragments thereof, and substantially similar sequence thereto. The screening can also be performed by selecting clones or any transgenic plants having a desired phenotype.

#### III.A.5. Regulatory Sequence Modulation

The SDFs described in Table 1, and fragments thereof are examples of nucleotides of the invention that contain regulatory sequences that can be used to suppress or inactivate transcription and/or translation from a gene of interest as discussed in I.C.5.

#### III.A.6. Genes Comprising Dominant-Negative Mutations

When suppression of production of the endogenous, native protein is desired it is often helpful to express a gene comprising a dominant negative mutation. Production of protein variants produced from genes comprising dominant negative mutations is a useful tool for research. Genes comprising dominant negative mutations can produce a variant polypeptide which is capable of competing with the native polypeptide, but which does not produce the native result. Consequently, over expression of genes comprising these mutations can titrate out an undesired activity of the native protein. For example, The product from a gene comprising a dominant negative mutation of a receptor can be used to constitutively activate or suppress a signal transduction cascade, allowing examination of the phenotype and thus the trait(s) controlled by that receptor and pathway. Alternatively, the protein arising from the gene comprising a dominant-negative mutation can be an inactive enzyme still capable of binding to the same substrate as the native protein and therefore competes with such native protein.

Products from genes comprising dominant-negative mutations can also act upon the native protein itself to prevent activity. For example, the native protein may be active only as a homo-multimer or as one subunit of a hetero-multimer. Incorporation of an inactive subunit into the multimer with native subunit(s) can inhibit activity.

Thus, gene function can be modulated in host cells of interest by insertion into these cells vector constructs comprising a gene comprising a dominant-negative mutation.

### III.B. Enhanced Expression

Enhanced expression of a gene of interest in a host cell can be accomplished by either (1) insertion of an exogenous gene; or (2) promoter modulation.

#### III.B.1. Insertion of an Exogenous Gene

Insertion of an expression construct encoding an exogenous gene can boost the number of gene copies expressed in a host cell.

Such expression constructs can comprise genes that either encode the native protein that is of interest or that encode a variant that exhibits enhanced activity as compared to the native protein. Such genes encoding proteins of interest can be constructed from the sequences from TABLE 1, fragments thereof, and substantially similar sequence thereto.

Such an exogenous gene can include either a constitutive promoter permitting expression in any cell in a host organism or a promoter that directs transcription only in particular cells or times during a host cell life cycle or in response to environmental stimuli.

#### III.B.2. Regulatory Sequence Modulation

The SDFs of Table 1, and fragments thereof, contain regulatory sequences that can be used to enhance expression of a gene of interest. For example, some of these sequences contain useful enhancer elements. In some cases, duplication of enhancer elements or insertion of exogenous enhancer elements will increase expression of a desired gene from a particular promoter. As other examples, all II promoters require binding of a regulatory protein to be activated, while some promoters may need a protein that signals a promoter binding protein to expose a polymerase binding site. In either case, over-production of such proteins can be used to enhance expression of a gene of interest by increasing the activation time of the promoter.

Such regulatory proteins are encoded by some of the sequences in TABLE 1, fragments thereof, and substantially similar sequences thereto.

Coding sequences for these proteins can be constructed as described above.

09679203-100400

#### IV. Gene Constructs and Vector Construction

To use isolated SDFs of the present invention or a combination of them or parts and/or mutants and/or fusions of said SDFs in the above techniques, recombinant DNA vectors which  
5 comprise said SDFs and are suitable for transformation of cells, such as plant cells, are usually prepared. The SDF construct can be made using standard recombinant DNA techniques (Sambrook et al. 1989) and can be introduced to the species of interest by *Agrobacterium*-mediated transformation or by other means of transformation (e.g., particle gun bombardment) as referenced below.

10 The vector backbone can be any of those typical in the art such as plasmids, viruses, artificial chromosomes, BACs, YACs and PACs and vectors of the sort described by

(a) **BAC:** Shizuya et al., Proc. Natl. Acad. Sci. USA 89: 8794-8797 (1992);  
Hamilton et al., Proc. Natl. Acad. Sci. USA 93: 9975-9979 (1996);

(b) **YAC:** Burke et al., Science 236:806-812 (1987);.

15 (c) **PAC:** Sternberg N. et al., Proc Natl Acad Sci U S A. Jan;87(1):103-7 (1990);

(d) **Bacteria-Yeast Shuttle Vectors:** Bradshaw et al., Nucl Acids Res 23: 4850-4856 (1995);

(e) **Lambda Phage Vectors:** Replacement Vector, e.g.,  
Frischauf et al., J. Mol Biol 170: 827-842 (1983); or Insertion vector, e.g.,

20 Huynh et al., In: Glover NM (ed) DNA Cloning: A practical Approach, Vol.1 Oxford: IRL Press (1985);

(f) **T-DNA gene fusion vectors :**Walden et al., Mol Cell Biol 1: 175-194 (1990);  
and

(g) **Plasmid vectors:** Sambrook et al., infra.

25 Typically, a vector will comprise the exogenous gene, which in its turn comprises an SDF of the present invention to be introduced into the genome of a host cell, and which gene may be an antisense construct, a ribozyme construct chimera, or a coding sequence with any desired transcriptional and/or translational regulatory sequences, such as promoters, UTRs, and 3' end termination sequences. Vectors of the invention can also include origins of  
30 replication, scaffold attachment regions (SARs), markers, homologous sequences, introns, etc.

A DNA sequence coding for the desired polypeptide, for example a cDNA sequence encoding a full length protein, will preferably be combined with transcriptional and translational

09679203-100400

initiation regulatory sequences which will direct the transcription of the sequence from the gene in the intended tissues of the transformed plant.

For example, for over-expression, a plant promoter fragment may be employed that will direct transcription of the gene in all tissues of a regenerated plant. Alternatively, the plant promoter may direct transcription of an SDF of the invention in a specific tissue (tissue-specific promoters) or may be otherwise under more precise environmental control (inducible promoters).

If proper polypeptide production is desired, a polyadenylation region at the 3'-end of the coding region is typically included. The polyadenylation region can be derived from the natural gene, from a variety of other plant genes, or from T-DNA.

The vector comprising the sequences from genes or SDF or the invention may comprise a marker gene that confers a selectable phenotype on plant cells. The vector can include promoter and coding sequence, for instance. For example, the marker may encode biocide resistance, particularly antibiotic resistance, such as resistance to kanamycin, G418, bleomycin, hygromycin, or herbicide resistance, such as resistance to chlorosulfuron or phosphinotricin.

#### IV.A. Coding Sequences

Generally, the sequence in the transformation vector and to be introduced into the genome of the host cell does not need to be absolutely identical to an SDF of the present invention. Also, it is not necessary for it to be full length, relative to either the primary transcription product or fully processed mRNA. Furthermore, the introduced sequence need not have the same intron or exon pattern as a native gene. Also, heterologous non-coding segments can be incorporated into the coding sequence without changing the desired amino acid sequence of the polypeptide to be produced.

#### IV.B. Promoters

As explained above, introducing an exogenous SDF from the same species or an orthologous SDF from another species can modulate the expression of a native gene corresponding to that SDF of interest. Such an SDF construct can be under the control of either a constitutive promoter or a highly regulated inducible promoter (*e.g.*, a copper inducible promoter). The promoter of interest can initially be either endogenous or heterologous to the species in question. When re-introduced into the genome of said species, such promoter becomes exogenous to said species. Over-expression of an SDF transgene can



lead to co-suppression of the homologous endogeneous sequence thereby creating some alterations in the phenotypes of the transformed species as demonstrated by similar analysis of the chalcone synthase gene (Napoli et al., *Plant Cell* 2:279 (1990) and van der Krol et al., *Plant Cell* 2:291 (1990)). If an SDF is found to encode a protein with desirable characteristics, its over-production can be controlled so that its accumulation can be manipulated in an organ- or tissue-specific manner utilizing a promoter having such specificity.

Likewise, if the promoter of an SDF (or an SDF that includes a promoter) is found to be tissue-specific or developmentally regulated, such a promoter can be utilized to drive or facilitate the transcription of a specific gene of interest (e.g., seed storage protein or root-specific protein). Thus, the level of accumulation of a particular protein can be manipulated or its spatial localization in an organ- or tissue- specific manner can be altered.

#### IV. C Signal Peptides

SDFs of the present invention containing signal peptides are indicated in Table 1. In some cases it may be desirable for the protein encoded by an introduced exogenous or orthologous SDF to be targeted (1) to a particular organelle intracellular compartment, (2) to interact with a particular molecule such as a membrane molecule or (3) for secretion outside of the cell harboring the introduced SDF. This will be accomplished using a signal peptide.

Signal peptides direct protein targeting, are involved in ligand-receptor interactions and act in cell to cell communication. Many proteins, especially soluble proteins, contain a signal peptide that targets the protein to one of several different intracellular compartments. In plants, these compartments include, but are not limited to, the endoplasmic reticulum (ER), mitochondria, plastids (such as chloroplasts), the vacuole, the Golgi apparatus, protein storage vessicles (PSV) and, in general, membranes. Some signal peptide sequences are conserved, such as the Asn-Pro-Ile-Arg amino acid motif found in the N-terminal propeptide signal that targets proteins to the vacuole (Marty (1999) *The Plant Cell* 11: 587-599). Other signal peptides do not have a consensus sequence *per se*, but are largely composed of hydrophobic amino acids, such as those signal peptides targeting proteins to the ER (Vitale and Denecke (1999) *The Plant Cell* 11: 615-628). Still others do not appear to contain either a consensus sequence or an identified common secondary sequence, for instance the chloroplast stromal targeting signal peptides (Keegstra and Cline (1999) *The Plant Cell* 11: 557-570). Furthermore, some targeting peptides are bipartite, directing proteins first to an organelle and then to a membrane within the organelle (e.g. within the thylakoid lumen of the

chloroplast; see Keegstra and Cline (1999) *The Plant Cell* 11: 557-570). In addition to the diversity in sequence and secondary structure, placement of the signal peptide is also varied. Proteins destined for the vacuole, for example, have targeting signal peptides found at the N-terminus, at the C-terminus and at a surface location in mature, folded proteins. Signal peptides also serve as ligands for some receptors.

These characteristics of signal proteins can be used to more tightly control the phenotypic expression of introduced SDFs. In particular, associating the appropriate signal sequence with a specific SDF can allow sequestering of the protein in specific organelles (plastids, as an example), secretion outside of the cell, targeting interaction with particular receptors, etc. Hence, the inclusion of signal proteins in constructs involving the SDFs of the invention increases the range of manipulation of SDF phenotypic expression. The nucleotide sequence of the signal peptide can be isolated from characterized genes using common molecular biological techniques or can be synthesized in vitro.

In addition, the native signal peptide sequences, both amino acid and nucleotide, described in Table 1 can be used to modulate polypeptide transport. Further variants of the native signal peptides described in Table 1 are contemplated. Insertions, deletions, or substitutions can be made. Such variants will retain at least one of the functions of the native signal peptide as well as exhibiting some degree of sequence identity to the native sequence.

Also, fragments of the signal peptides of the invention are useful and can be fused with other signal peptides of interest to modulate transport of a polypeptide.

## V. Transformation Techniques

A wide range of techniques for inserting exogenous polynucleotides are known for a number of host cells, including, without limitation, bacterial, yeast, mammalian, insect and plant cells.

Techniques for transforming a wide variety of higher plant species are well known and described in the technical and scientific literature. See, e.g. Weising et al., *Ann. Rev. Genet.* 22:421 (1988); and Christou, *Euphytica*, v. 85, n.1-3:13-27, (1995).

DNA constructs of the invention may be introduced into the genome of the desired plant host by a variety of conventional techniques. For example, the DNA construct may be introduced directly into the genomic DNA of the plant cell using techniques such as electroporation and microinjection of plant cell protoplasts, or the DNA constructs can be introduced directly to plant tissue using ballistic methods, such as DNA particle bombardment. Alternatively, the DNA constructs may be combined with suitable T-DNA flanking regions and

introduced into a conventional *Agrobacterium tumefaciens* host vector. The virulence functions of the *Agrobacterium tumefaciens* host will direct the insertion of the construct and adjacent marker into the plant cell DNA when the cell is infected by the bacteria (McCormac et al., *Mol. Biotechnol.* 8:199 (1997); Hamilton, *Gene* 200:107 (1997)); Salomon et al. *EMBO J.* 3:141 (1984); Herrera-Estrella et al. *EMBO J.* 2:987 (1983).

Microinjection techniques are known in the art and well described in the scientific and patent literature. The introduction of DNA constructs using polyethylene glycol precipitation is described in Paszkowski et al. *EMBO J.* 3:2717 (1984). Electroporation techniques are described in Fromm et al. *Proc. Natl Acad. Sci. USA* 82:5824 (1985). Ballistic transformation techniques are described in Klein et al. *Nature* 327:773 (1987). *Agrobacterium tumefaciens*-mediated transformation techniques, including disarming and use of binary or co-integrate vectors, are well described in the scientific literature. See, for example Hamilton, CM., *Gene* 200:107 (1997); Müller et al. *Mol. Gen. Genet.* 207:171 (1987); Komari et al. *Plant J.* 10:165 (1996); Venkateswarlu et al. *Biotechnology* 9:1103 (1991) and Gleave, AP., *Plant Mol. Biol.* 20:1203 (1992); Graves and Goldman, *Plant Mol. Biol.* 7:34 (1986) and Gould et al., *Plant Physiology* 95:426 (1991).

Transformed plant cells which are derived by any of the above transformation techniques can be cultured to regenerate a whole plant that possesses the transformed genotype and thus the desired phenotype such as seedlessness. Such regeneration techniques rely on manipulation of certain phytohormones in a tissue culture growth medium, typically relying on a biocide and/or herbicide marker which has been introduced together with the desired nucleotide sequences. Plant regeneration from cultured protoplasts is described in Evans et al., *Protoplasts Isolation and Culture* in "Handbook of Plant Cell Culture," pp. 124-176, MacMillan Publishing Company, New York, 1983; and Binding, *Regeneration of Plants, Plant Protoplasts*, pp. 21-73, CRC Press, Boca Raton, 1988. Regeneration can also be obtained from plant callus, explants, organs, or parts thereof. Such regeneration techniques are described generally in Klee et al. *Ann. Rev. of Plant Phys.* 38:467 (1987). Regeneration of monocots (rice) is described by Hosoyama et al. (*Biosci. Biotechnol. Biochem.* 58:1500 (1994)) and by Ghosh et al. (*J. Biotechnol.* 32:1 (1994)). The nucleic acids of the invention can be used to confer desired traits on essentially any plant.

Thus, the invention has use over a broad range of plants, including species from the genera *Anacardium*, *Arachis*, *Asparagus*, *Atropa*, *Avena*, *Brassica*, *Citrus*, *Citrullus*, *Capsicum*, *Carthamus*, *Cocos*, *Coffea*, *Cucumis*, *Cucurbita*, *Daucus*, *Elaeis*, *Fragaria*, *Glycine*, *Gossypium*, *Helianthus*, *Heterocallis*, *Hordeum*, *Hyoscyamus*, *Lactuca*, *Linum*, *Lolium*, *Lupinus*,

*Lycopersicon, Malus, Manihot, Majorana, Medicago, Nicotiana, Olea, Oryza, Panieum, Pannesetum, Persea, Phaseolus, Pistachia, Pisum, Pyrus, Prunus, Raphanus, Ricinus, Secale, Senecio, Sinapis, Solanum, Sorghum, Theobromus, Trigonella, Triticum, Vicia, Vitis, Vigna, and, Zea.*

5 One of skill will recognize that after the expression cassette is stably incorporated in transgenic plants and confirmed to be operable, it can be introduced into other plants by sexual crossing. Any of a number of standard breeding techniques can be used, depending upon the species to be crossed.

The particular sequences of SDFs identified are provided in the attached TABLE 1.

10 One of ordinary skill in the art, having this data, can obtain cloned DNA fragments, synthetic DNA fragments or polypeptides constituting desired sequences by recombinant methodology known in the art or described herein.

## EXAMPLES

The invention is illustrated by way of the following examples. The invention is not  
15 limited by these examples as the scope of the invention is defined solely by the claims following.

### EXAMPLE 1: cDNA PREPARATION

A number of the nucleotide sequences disclosed in TABLE 1 herein as representative of the SDFs of the invention can be obtained by sequencing genomic DNA (gDNA) and/or cDNA  
20 from corn plants grown from HYBRID SEED # 35A19, purchased from Pioneer Hi-Bred International, Inc., Supply Management, P.O. Box 256, Johnston, Iowa 50131-0256.

A number of the nucleotide sequences disclosed in TABLE 1 herein as representative of the SDFs of the invention can also be obtained by sequencing genomic DNA from  
25 *Arabidopsis thaliana*, Wassilewskija ecotype or by sequencing cDNA obtained from mRNA from such plants as described below. This is a true breeding strain. Seeds of the plant are available from the Arabidopsis Biological Resource Center at the Ohio State University, under the accession number CS2360. Seeds of this plant were deposited under the terms and conditions of the Budapest Treaty at the American Type Culture Collection, Manassas, VA on August 31, 1999, and were assigned ATCC No. PTA-595.

30 Other methods for cloning full-length cDNA are described, for example, by Seki et al., *Plant Journal* **15**:707-720 (1998) High-efficiency cloning of Arabidopsis full-length cDNA by biotinylated Cap trapper"; Maruyama et al., *Gene* **138**:171 (1994) Oligo-capping a

09679203 100400

simple method to replace the cap structure of eukaryotic mRNAs with oligoribonucleotides"; and WO 96/34981.

Tissues were, or each organ was, individually pulverized and frozen in liquid nitrogen. Next, the samples were homogenized in the presence of detergents and then  
5 centrifuged. The debris and nuclei were removed from the sample and more detergents were added to the sample. The sample was centrifuged and the debris was removed. Then the sample was applied to a 2M sucrose cushion to isolate polysomes. The RNA was isolated by treatment with detergents and proteinase K followed by ethanol precipitation and centrifugation. The polysomal RNA from the different tissues was pooled according to the  
10 following mass ratios: 15/15/1 for male inflorescences, female inflorescences and root, respectively. The pooled material was then used for cDNA synthesis by the methods described below.

Starting material for cDNA synthesis for the exemplary corn cDNA clones with sequences presented in TABLE 1 was poly(A)-containing polysomal mRNAs from  
15 inflorescences and root tissues of corn plants grown from HYBRID SEED # 35A19. Male inflorescences and female (pre-and post-fertilization) inflorescences were isolated at various stages of development. Selection for poly(A) containing polysomal RNA was done using oligo d(T) cellulose columns, as described by Cox and Goldberg, Plant Molecular Biology: A Practical Approach", pp. 1-35, Shaw ed., c. 1988 by IRL, Oxford. The quality and the  
20 integrity of the polyA+ RNAs were evaluated.

Starting material for cDNA synthesis for the exemplary *Arabidopsis* cDNA clones with sequences presented in TABLE 1 was polysomal RNA isolated from the top-most inflorescence tissues of *Arabidopsis thaliana* Wassilewskija (Ws.) and from roots of  
25 *Arabidopsis thaliana* Landsberg erecta (L. er.), also obtained from the Arabidopsis Biological Resource Center. Nine parts inflorescence to every part root was used, as measured by wet mass. Tissue was pulverized and exposed to liquid nitrogen. Next, the sample was homogenized in the presence of detergents and then centrifuged. The debris and nuclei were removed from the sample and more detergents were added to the sample. The  
30 sample was centrifuged and the debris was removed and the sample was applied to a 2M sucrose cushion to isolate polysomal RNA. Cox et al., Plant Molecular Biology: A Practical Approach", pp. 1-35, Shaw ed., c. 1988 by IRL, Oxford. The polysomal RNA was used for cDNA synthesis by the methods described below. Polysomal mRNA was then isolated as described above for corn cDNA. The quality of the RNA was assessed electrophoretically.

09679203-100400

Following preparation of the mRNAs from various tissues as described above, selection of mRNA with intact 5' ends and specific attachment of an oligonucleotide tag to the 5' end of such mRNA was performed using either a chemical or enzymatic approach. Both techniques take advantage of the presence of the "cap" structure, which characterizes the 5' end of most intact mRNAs and which comprises a guanosine generally methylated once, at the 7 position.

The chemical modification approach involves the optional elimination of the 2', 3'-cis diol of the 3' terminal ribose, the oxidation of the 2', 3'-cis diol of the ribose linked to the cap of the 5' ends of the mRNAs into a dialdehyde, and the coupling of the such obtained dialdehyde to a derivatized oligonucleotide tag. Further detail regarding the chemical approaches for obtaining mRNAs having intact 5' ends are disclosed in International Application No. WO96/34981 published November 7, 1996.

The enzymatic approach for ligating the oligonucleotide tag to the intact 5' ends of mRNAs involves the removal of the phosphate groups present on the 5' ends of uncapped incomplete mRNAs, the subsequent decapping of mRNAs having intact 5' ends and the ligation of the phosphate present at the 5' end of the decapped mRNA to an oligonucleotide tag. Further detail regarding the enzymatic approaches for obtaining mRNAs having intact 5' ends are disclosed in Dumas Milne Edwards J.B. (Doctoral Thesis of Paris VI University, *Le clonage des ADNc complets: difficultés et perspectives nouvelles. Apports pour l'étude de la régulation de l'expression de la tryptophane hydroxylase de rat*, 20 Dec. 1993), EP0 625572 and Kato *et al.*, *Gene* 150:243-250 (1994).

In both the chemical and the enzymatic approach, the oligonucleotide tag has a restriction enzyme site (e.g. an EcoRI site) therein to facilitate later cloning procedures. Following attachment of the oligonucleotide tag to the mRNA, the integrity of the mRNA is examined by performing a Northern blot using a probe complementary to the oligonucleotide tag.

For the mRNAs joined to oligonucleotide tags using either the chemical or the enzymatic method, first strand cDNA synthesis is performed using an oligo-dT primer with reverse transcriptase. This oligo-dT primer can contain an internal tag of at least 4 nucleotides, which can be different from one mRNA preparation to another. Methylated dCTP is used for cDNA first strand synthesis to protect the internal EcoRI sites from digestion during subsequent steps. The first strand cDNA is precipitated using isopropanol after removal of RNA by alkaline hydrolysis to eliminate residual primers.

Second strand cDNA synthesis is conducted using a DNA polymerase, such as Klenow fragment and a primer corresponding to the 5' end of the ligated oligonucleotide. The primer is

004001" E0262960

typically 20-25 bases in length. Methylated dCTP is used for second strand synthesis in order to protect internal EcoRI sites in the cDNA from digestion during the cloning process.

Following second strand synthesis, the full-length cDNAs are cloned into a phagemid vector, such as pBlueScript™ (Stratagene). The ends of the full-length cDNAs are blunted with  
5 T4 DNA polymerase (Biolabs) and the cDNA is digested with EcoRI. Since methylated dCTP is used during cDNA synthesis, the EcoRI site present in the tag is the only hemi-methylated site; hence the only site susceptible to EcoRI digestion. In some instances, to facilitate subcloning, an Hind III adapter is added to the 3' end of full-length cDNAs.

The full-length cDNAs are then size fractionated using either exclusion chromatography  
10 (AcA, Biosepra) or electrophoretic separation which yields 3 to 6 different fractions. The full-length cDNAs are then directionally cloned either into pBlueScript™ using either the EcoRI and SmaI restriction sites or, when the Hind III adapter is present in the full-length cDNAs, the EcoRI and Hind III restriction sites. The ligation mixture is transformed, preferably by electroporation, into bacteria, which are then propagated under appropriate antibiotic selection.

15 Clones containing the oligonucleotide tag attached to full-length cDNAs are selected as follows.

The plasmid cDNA libraries made as described above are purified (e.g. by a column available from Qiagen). A positive selection of the tagged clones is performed as follows. Briefly, in this selection procedure, the plasmid DNA is converted to single stranded DNA using  
20 phage F1 gene II endonuclease in combination with an exonuclease (Chang et al., *Gene* 127:95 (1993)) such as exonuclease III or T7 gene 6 exonuclease. The resulting single stranded DNA is then purified using paramagnetic beads as described by Fry et al., *Biotechniques* 13: 124 (1992). Here the single stranded DNA is hybridized with a biotinylated oligonucleotide having a sequence corresponding to the 3' end of the oligonucleotide tag. Preferably, the primer has a  
25 length of 20-25 bases. Clones including a sequence complementary to the biotinylated oligonucleotide are selected by incubation with streptavidin coated magnetic beads followed by magnetic capture. After capture of the positive clones, the plasmid DNA is released from the magnetic beads and converted into double stranded DNA using a DNA polymerase such as ThermoSequenase™ (obtained from Amersham Pharmacia Biotech). Alternatively, protocols  
30 such as the Gene Trapper™ kit (Gibco BRL) can be used. The double stranded DNA is then transformed, preferably by electroporation, into bacteria. The percentage of positive clones having the 5' tag oligonucleotide is typically estimated to be between 90 and 98% from dot blot analysis.

004001" E0262960

Following transformation, the libraries are ordered in microtiter plates and sequenced. The *Arabidopsis* library was deposited at the American Type Culture Collection on January 7, 2000 as *E-coli* liba 010600" under the accession number **PTA-1161**.

## **EXAMPLE 2: SOUTHERN HYBRIDIZATIONS**

5 The SDFs of the invention can be used in Southern hybridizations as described above. The following describes extraction of DNA from nuclei of plant cells, digestion of the nuclear DNA and separation by length, transfer of the separated fragments to membranes, preparation of probes for hybridization, hybridization and detection of the hybridized probe.

The procedures described herein can be used to isolate related polynucleotides or for  
10 diagnostic purposes. Moderate stringency hybridization conditions, as defined above, are described in the present example. These conditions result in detection of hybridization between sequences having at least 70% sequence identity. As described above, the hybridization and wash conditions can be changed to reflect the desired percentatge of sequence identity between probe and target sequences that can be detected.

15 In the following procedure, a probe for hybridization is produced from two PCR reactions using two primers from genomic sequence of *Arabidopsis thaliana*. As described above, the particular template for generating the probe can be any desired template.

The first PCR product is assessed to validate the size of the primer to assure it is of the expected size. Then the product of the first PCR is used as a template, with the same pair  
20 of primers used in the first PCR, in a second PCR that produces a labeled product used as the probe.

Fragments detected by hybridization, or other bands of interest, can be isolated from gels used to separate genomic DNA fragments by known methods for further purification and/or characterization.

## 25 **Buffers for nuclear DNA extraction**

### 1. 10X HB

|                  | 1000 ml |   |
|------------------|---------|---|
| 40 mM spermidine | 10.2 g  | Spermine (Sigma S-2876) and spermidine (Sigma S-2501) |
| 10 mM spermine   | 3.5 g   | Stabilize chromatin and the nuclear membrane          |



|                          |        |  |
|--------------------------|--------|--|
| 0.1 M EDTA<br>(disodium) | 37.2 g | EDTA inhibits nuclease                         |
| 0.1 M Tris               | 12.1 g | Buffer   |
| 0.8 M KCl                | 59.6 g | Adjusts ionic strength for stability of nuclei |

Adjust pH to 9.5 with 10 N NaOH. It appears that there is a nuclease present in leaves. Use of pH 9.5 appears to inactivate this nuclease.

2. 2 M sucrose (684 g per 1000 ml)

Heat about half the final volume of water to about 50°C. Add the sucrose slowly then bring the mixture to close to final volume; stir constantly until it has dissolved. Bring the solution to volume.

3. Sarkosyl solution (lyses nuclear membranes)

1000 ml

N-lauroyl sarcosine (Sarkosyl) 20.0 g

0.1 M Tris 12.1 g

0.04 M EDTA (Disodium) 14.9 g

Adjust the pH to 9.5 after all the components are dissolved and bring up to the proper volume.

4. 20% Triton X-100

80 ml Triton X-100

320 ml 1xHB (w/o  $\beta$ -ME and PMSF)

Prepare in advance; Triton takes some time to dissolve

A. Procedure

1. Prepare 1X H<sup>+</sup> buffer (keep ice-cold during use)

004007 E0262960

|             |                              |
|-------------|------------------------------|
|             | <u>1000 ml</u>               |
| 10X HB      | 100 ml                       |
| 2 M sucrose | 250 ml a non-ionic osmoticum |
| Water       | 634 ml                       |

5 **Added just before use:**

|                          |   |
|--------------------------|---|
| 100 mM PMSF*             | 10 ml a protease inhibitor; protects<br>nuclear membrane proteins |
| $\beta$ -mercaptoethanol | 1 ml inactivates nuclease by reducing<br>disulfide bonds          |

10 \*100 mM PMSF  
(phenyl methyl sulfonyl fluoride, Sigma P-7626)  
(add 0.0875 g to 5 ml 100% ethanol)

15 2. Homogenize the tissue in a blender (use 300-400 ml of 1xHB per blender). Be sure that you use 5-10 ml of HB buffer per gram of tissue. Blenders generate heat so be sure to keep the homogenate cold. It is necessary to put the blenders in ice periodically.

3. Add the 20% Triton X-100 (25 ml per liter of homogenate) and gently stir on ice for 20 min. This lyses plastid, but not nuclear, membranes.

20 4. Filter the tissue suspension through several nylon filters into an ice-cold beaker. The first filtration is through a 250-micron membrane; the second is through an 85-micron membrane; the third is through a 50-micron membrane; and the fourth is through a 20-micron membrane. Use a large funnel to hold the filters. Filtration can be sped up by gently squeezing the liquid through the filters.

5. Centrifuge the filtrate at 1200 x g for 20 min. at 4°C to pellet the nuclei.

004007 00264960

6. Discard the dark green supernatant. The pellet will have several layers to it. One is starch; it is white and gritty. The nuclei are gray and soft. In the early steps, there may be a dark green and somewhat viscous layer of chloroplasts.

Wash the pellets in about 25 ml cold H buffer (with Triton X-100) and resuspend by swirling gently and pipetting. After the pellets are resuspended.

Pellet the nuclei again at 1200 - 1300 x g. Discard the supernatant.

Repeat the wash 3-4 times until the supernatant has changed from a dark green to a pale green. This usually happens after 3 or 4 resuspensions. At this point, the pellet is typically grayish white and very slippery. The Triton X-100 in these repeated steps helps to destroy the chloroplasts and mitochondria that contaminate the prep.

Resuspend the nuclei for a final time in a total of 15 ml of H buffer and transfer the suspension to a sterile 125 ml Erlenmeyer flask.

7. Add 15 ml, dropwise, cold 2% Sarkosyl, 0.1 M Tris, 0.04 M EDTA solution (pH 9.5) while swirling gently. This lyses the nuclei. The solution will become very viscous.

8. Add 30 grams of CsCl and gently swirl at room temperature until the CsCl is in solution. The mixture will be gray, white and viscous.

9. Centrifuge the solution at 11,400 x g at 4°C for at least 30 min. The longer this spin is, the firmer the protein pellicle.

10. The result is typically a clear green supernatant over a white pellet, and (perhaps) under a protein pellicle. Carefully remove the solution under the protein pellicle and above the pellet. Determine the density of the solution by weighing 1 ml of solution and add CsCl if necessary to bring to 1.57 g/ml. The solution contains dissolved solids (sucrose etc) and the refractive index alone will not be an accurate guide to CsCl concentration.

11. Add 20  $\mu$ l of 10 mg/ml EtBr per ml of solution.
12. Centrifuge at 184,000 x g for 16 to 20 hours in a fixed-angle rotor.
13. Remove the dark red supernatant that is at the top of the tube with a plastic transfer pipette and discard. Carefully remove the DNA band with another transfer pipette.  
5 The DNA band is usually visible in room light; otherwise, use a long wave UV light to locate the band.
14. Extract the ethidium bromide with isopropanol saturated with water and salt. Once the solution is clear, extract at least two more times to ensure that all of the EtBr is gone. Be very gentle, as it is very easy to shear the DNA at this step. This extraction may take a while because the DNA solution tends to be very viscous. If the solution is too viscous, dilute it with TE.  
10
15. Dialyze the DNA for at least two days against several changes (at least three times) of TE (10 mM Tris, 1mM EDTA, pH 8) to remove the cesium chloride.
16. Remove the dialyzed DNA from the tubing. If the dialyzed DNA solution contains a lot of debris, centrifuge the DNA solution at least at 2500 x g for 10 min. and carefully transfer the clear supernatant to a new tube. Read the A260 concentration of the DNA.  
15
17. Assess the quality of the DNA by agarose gel electrophoresis (1% agarose gel) of the DNA. Load 50 ng and 100 ng (based on the OD reading) and compare it with known and good quality DNA. Undigested lambda DNA and a lambda-HindIII-digested DNA are good molecular weight makers.  
20

### Protocol for Digestion of Genomic DNA

#### Protocol:

1. The relative amounts of DNA for different crop plants that provide approximately a balanced number of genome equivalent is given in Table 3. Note that due to the size

004001" 00264960

of the wheat genome, wheat DNA will be underrepresented. Lambda DNA provides a useful control for complete digestion.

2. Precipitate the DNA by adding 3 volumes of 100% ethanol. Incubate at  $-20^{\circ}\text{C}$  for at least two hours. Yeast DNA can be purchased and made up at the necessary concentration, therefore no precipitation is necessary for yeast DNA.

3. Centrifuge the solution at  $11,400 \times g$  for 20 min. Decant the ethanol carefully (be careful not to disturb the pellet). Be sure that the residual ethanol is completely removed either by vacuum desiccation or by carefully wiping the sides of the tubes with a clean tissue.

4. Resuspend the pellet in an appropriate volume of water. Be sure the pellet is fully resuspended before proceeding to the next step. This may take about 30 min.

5. Add the appropriate volume of 10X reaction buffer provided by the manufacturer of the restriction enzyme to the resuspended DNA followed by the appropriate volume of enzymes. Be sure to mix it properly by slowly swirling the tubes.

6. Set-up the lambda digestion-control for each DNA that you are digesting.

7. Incubate both the experimental and lambda digests overnight at  $37^{\circ}\text{C}$ . Spin down condensation in a microfuge before proceeding.

8. After digestion, add  $2 \mu\text{l}$  of loading dye (typically 0.25% bromophenol blue, 0.25% xylene cyanol in 15% Ficoll or 30% glycerol) to the lambda-control digests and load in 1% TPE-agarose gel (TPE is 90 mM Tris-phosphate, 2 mM EDTA, pH 8). If the lambda DNA in the lambda control digests are completely digested, proceed with the precipitation of the genomic DNA in the digests.

9. Precipitate the digested DNA by adding 3 volumes of 100% ethanol and incubating in  $-20^{\circ}\text{C}$  for at least 2 hours (preferably overnight).

09679203 100400

EXCEPTION: *Arabidopsis* and yeast DNA are digested in an appropriate volume; they don't have to be precipitated.

10. Resuspend the DNA in an appropriate volume of TE (e.g., 22  $\mu$ l x 50 blots = 1100  $\mu$ l) and an appropriate volume of 10X loading dye (e.g., 2.4  $\mu$ l x 50 blots = 120  $\mu$ l). Be careful in pipetting the loading dye - it is viscous. Be sure you are pipetting the correct volume.

Table 3

Some guide points in digesting genomic DNA.

| Species            | Genome Size | Size Relative to <i>Arabidopsis</i> | Genome Equivalent to 2 $\mu$ g <i>Arabidopsis</i> DNA | Amount of DNA per blot |
|--------------------|-------------|-------------------------------------|---|------------------------|
| <i>Arabidopsis</i> | 120 Mb      | 1X                                  | 1X  | 2 $\mu$ g              |
| Brassica           | 1,100 Mb    | 9.2X                                | 0.54X   | 10 $\mu$ g             |
| Corn               | 2,800 Mb    | 23.3X                               | 0.43X   | 20 $\mu$ g             |
| Cotton             | 2,300 Mb    | 19.2X                               | 0.52X   | 20 $\mu$ g             |
| Oat                | 11,300 Mb   | 94X                                 | 0.11X   | 20 $\mu$ g             |
| Rice               | 400 Mb      | 3.3X                                | 0.75X   | 5 $\mu$ g              |
| Soybean            | 1,100 Mb    | 9.2X                                | 0.54X   | 10 $\mu$ g             |
| Sugarbeet          | 758 Mb      | 6.3X                                | 0.8X  | 10 $\mu$ g             |
| Sweetclover        | 1,100 Mb    | 9.2X                                | 0.54X   | 10 $\mu$ g             |
| Wheat              | 16,000 Mb   | 133X                                | 0.08X   | 20 $\mu$ g             |
| Yeast              | 15 Mb       | 0.12X                               | 1X  | 0.25 $\mu$ g           |

### Protocol for Southern Blot Analysis

The digested DNA samples are electrophoresed in 1% agarose gels in 1x TPE buffer. Low voltage; overnight separations are preferred. The gels are stained with EtBr and photographed.

1. For blotting the gels, first incubate the gel in 0.25 N HCl (with gentle shaking) for about 15 min.
2. Then briefly rinse with water. The DNA is denatured by 2 incubations. Incubate (with shaking) in 0.5 M NaOH in 1.5 M NaCl for 15 min.
3. The gel is then briefly rinsed in water and neutralized by incubating twice (with shaking) in 1.5 M Tris pH 7.5 in 1.5 M NaCl for 15 min.
4. A nylon membrane is prepared by soaking it in water for at least 5 min, then in 6X SSC for at least 15 min. before use. (20x SSC is 175.3 g NaCl, 88.2 g sodium citrate per liter, adjusted to pH 7.0.)
5. The nylon membrane is placed on top of the gel and all bubbles in between are removed. The DNA is blotted from the gel to the membrane using an absorbent medium, such as paper toweling and 6x SCC buffer. After the transfer, the membrane may be lightly brushed with a gloved hand to remove any agarose sticking to the surface.
6. The DNA is then fixed to the membrane by UV crosslinking and baking at 80°C. The membrane is stored at 4°C until use.

**B. Protocol for PCR Amplification of Genomic Fragments in Arabidopsis**

**Amplification procedures:**

1. Mix the following in a 0.20 ml PCR tube or 96-well PCR plate:

| Volume | Stock                               | Final Amount or Conc. |
|--------|-------------------------------------|-----------------------|
| 0.5 µl | ~ 10 ng/µl genomic DNA <sup>1</sup> | 5 ng                  |
| 2.5 µl | <b>10X PCR buffer</b>               | 20 mM Tris, 50 mM KCl |

<sup>1</sup> Arabidopsis DNA is used in the present experiment, but the procedure is a general one.

|                 |  |         |
|-----------------|--|---------|
| 0.75 $\mu$ l    | 50 mM MgCl <sub>2</sub>  | 1.5 mM  |
| 1 $\mu$ l       | 10 pmol/ $\mu$ l Primer 1 (Forward)  | 10 pmol |
| 1 $\mu$ l       | 10 pmol/ $\mu$ l Primer 2 (Reverse)  | 10 pmol |
| 0.5 $\mu$ l     | 5 mM dNTPs   | 0.1 mM  |
| 0.1 $\mu$ l     | 5 units/ $\mu$ l Platinum Taq™ (Life Technologies, Gaithersburg, MD)<br>DNA Polymerase | 1 units |
| (to 25 $\mu$ l) | <b>Water</b>   |         |

2. The template DNA is amplified using a Perkin Elmer 9700 PCR machine:

1) 94°C for 10 min. followed by

| 2)<br>5 cycles: | 3)<br>5 cycles: | 4)<br>25 cycles: |
|-----------------|-----------------|------------------|
| 94 °C - 30 sec  | 94 °C - 30 sec  | 94 °C - 30 sec   |
| 62 °C - 30 sec  | 58 °C - 30 sec  | 53 °C - 30 sec   |
| 72 °C - 3 min   | 72 °C - 3 min   | 72 °C - 3 min    |

5) 72°C for 7 min. Then the reactions are stopped by chilling to 4°C.

The procedure can be adapted to a multi-well format if necessary.

## 5 Quantification and Dilution of PCR Products:

1. The product of the PCR is analyzed by electrophoresis in a 1% agarose gel. A linearized plasmid DNA can be used as a quantification standard (usually at 50, 100,

004007" 00264960



200, and 400 ng). These will be used as references to approximate the amount of PCR products. HindIII-digested Lambda DNA is useful as a molecular weight marker. The gel can be run fairly quickly; e.g., at 100 volts. The standard gel is examined to determine that the size of the PCR products is consistent with the expected size and if there are significant extra bands or smeary products in the PCR reactions.

2. The amounts of PCR products can be estimated on the basis of the plasmid standard.
3. For the small number of reactions that produce extraneous bands, a small amount of DNA from bands with the correct size can be isolated by dipping a sterile 10- $\mu$ l tip into the band while viewing through a UV Transilluminator. The small amount of agarose gel (with the DNA fragment) is used in the labeling reaction.

### C. Protocol for PCR-DIG-Labeling of DNA

#### Solutions:

Reagents in PCR reactions (diluted PCR products, 10X PCR Buffer, 50 mM MgCl<sub>2</sub>, 5 U/ $\mu$ l Platinum Taq Polymerase, and the primers)

10X dNTP + DIG-11-dUTP [1:5]: (2 mM dATP, 2 mM dCTP, 2 mM dGTP, 1.65 mM dTTP, 0.35 mM DIG-11-dUTP)

10X dNTP + DIG-11-dUTP [1:10]: (2 mM dATP, 2 mM dCTP, 2 mM dGTP, 1.81 mM dTTP, 0.19 mM DIG-11-dUTP)

10X dNTP + DIG-11-dUTP [1:15]: (2 mM dATP, 2 mM dCTP, 2 mM dGTP, 1.875 mM dTTP, 0.125 mM DIG-11-dUTP)

TE buffer (10 mM Tris, 1 mM EDTA, pH 8)

Maleate buffer: In 700 ml of deionized distilled water, dissolve 11.61 g maleic acid and 8.77 g NaCl. Add NaOH to adjust the pH to 7.5. Bring the volume to 1 L. Stir for 15 min. and sterilize.

10% blocking solution: In 80 ml deionized distilled water, dissolve 1.16g maleic acid. Next, add NaOH to adjust the pH to 7.5. Add 10 g of the blocking reagent powder (Boehringer Mannheim, Indianapolis, IN, Cat. no. 1096176). Heat to 60°C while stirring to dissolve the powder. Adjust the volume to 100 ml with water. Stir and sterilize.

5

1% blocking solution: Dilute the 10% stock to 1% using the maleate buffer.

Buffer 3 (100 mM Tris, 100 mM NaCl, 50 mM MgCl<sub>2</sub>, pH9.5). Prepared from autoclaved solutions of 1M Tris pH 9.5, 5 M NaCl, and 1 M MgCl<sub>2</sub> in autoclaved distilled water.

09679203 100400

Procedure:

1. PCR reactions are performed in 25 µl volumes containing:

|                          |                                |
|--------------------------|--------------------------------|
| PCR buffer               | 1X                             |
| MgCl <sub>2</sub>        | 1.5 mM                         |
| 10X dNTP + DIG-11-dUTP   | 1X (please see the note below) |
| Platinum Taq™ Polymerase | 1 unit                         |
| 10 pg probe DNA          |                                |
| 10 pmol primer 1         |                                |

Note:Use for:

|                               |                |
|-------------------------------|----------------|
| 10X dNTP + DIG-11-dUTP (1:5)  | < 1 kb         |
| 10X dNTP + DIG-11-dUTP (1:10) | 1 kb to 1.8 kb |
| 10X dNTP + DIG-11-dUTP (1:15) | > 1.8 kb       |

2. The PCR reaction uses the following amplification cycles:

- 1) 94°C for 10 min.

| 2)<br>5 cycles: | 3)<br>5 cycles: | 4)<br>25 cycles: |
|-----------------|-----------------|------------------|
| 95°C - 30 sec   | 95°C - 30 sec   | 95°C - 30 sec    |
| 61°C - 1 min    | 59°C - 1 min    | 51°C - 1 min     |
| 73°C - 5 min    | 75°C - 5 min    | 73°C - 5 min     |

- 5) 72°C for 8 min. The reactions are terminated by chilling to 4°C (hold).

3. The products are analyzed by electrophoresis- in a 1% agarose gel, comparing to an aliquot of the unlabelled probe starting material.

4. The amount of DIG-labeled probe is determined as follows:

Make serial dilutions of the diluted control DNA in dilution buffer (TE: 10 mM Tris and 1 mM EDTA, pH 8) as shown in the following table:

| DIG-labeled control<br>DNA starting conc. | Stepwise Dilution           | Final Conc. (Dilution<br>Name) |
|---|-----------------------------|--------------------------------|
| 5 ng/ $\mu$ l                             | 1 $\mu$ l in 49 $\mu$ l TE  | 100 pg/ $\mu$ l (A)            |
| 100 pg/ $\mu$ l (A)                       | 25 $\mu$ l in 25 $\mu$ l TE | 50 pg/ $\mu$ l (B)             |
| 50 pg/ $\mu$ l (B)                        | 25 $\mu$ l in 25 $\mu$ l TE | 25 pg/ $\mu$ l (C)             |
| 25 pg/ $\mu$ l (C)                        | 20 $\mu$ l in 30 $\mu$ l TE | 10 pg/ $\mu$ l (D)             |

- a. Serial dilutions of a DIG-labeled standard DNA ranging from 100 pg to 10 pg are spotted onto a positively charged nylon membrane, marking the membrane lightly with a pencil to identify each dilution.
- b. Serial dilutions (e.g., 1:50, 1:2500, 1:10,000) of the newly labeled DNA probe are spotted.
- c. The membrane is fixed by UV crosslinking.
- d. The membrane is wetted with a small amount of maleate buffer and then incubated in 1% blocking solution for 15 min at room temp.
- e. The labeled DNA is then detected using alkaline phosphatase conjugated anti-DIG antibody (Boehringer Mannheim, Indianapolis, IN, cat. no. 1093274) and an NBT substrate according to the manufacture's instruction.
- f. Spot intensities of the control and experimental dilutions are then compared to estimate the concentration of the PCR-DIG-labeled probe.

**D. Prehybridization and Hybridization of Southern Blots**Solutions:

100% Formamide purchased from Gibco

20X SSC (1X = 0.15 M NaCl, 0.015 M Na<sub>3</sub>citrate)

per L: 175 g NaCl  
87.5 g Na<sub>3</sub>citrate·2H<sub>2</sub>O

20% Sarkosyl (N-lauroyl-sarcosine)

20% SDS (sodium dodecyl sulphate)

10% Blocking Reagent: In 80 ml deionized distilled water, dissolve 1.16 g maleic acid. Next, add NaOH to adjust the pH to 7.5. Add 10 g of the blocking reagent powder. Heat to 60°C while stirring to dissolve the powder. Adjust the volume to 100 ml with water. Stir and sterilize.

Prehybridization Mix:

| Final Concentration | Components       | Volume (per 100 ml) | Stock |
|---------------------|------------------|---------------------|-------|
| 50%                 | Formamide        | 50 ml               | 100%  |
| 5X                  | SSC              | 25 ml               | 20X   |
| 0.1%                | Sarkosyl         | 0.5 ml              | 20%   |
| 0.02%               | SDS              | 0.1 ml              | 20%   |
| 2%                  | Blocking Reagent | 20 ml               | 10%   |
|                     | Water            | 4.4 ml              |       |

General Procedures:

- Place the blot in a heat-sealable plastic bag and add an appropriate volume of prehybridization solution (30 ml/100cm<sup>2</sup>) at room temperature. Seal the bag with a heat sealer, avoiding bubbles as much as possible. Lay down the bags in a large plastic tray (one tray can accommodate at least 4-5 bags). Ensure that the bags are

lying flat in the tray so that the prehybridization solution is evenly distributed throughout the bag. Incubate the blot for at least 2 hours with gentle agitation using a waver shaker.

2. Denature DIG-labeled DNA probe by incubating for 10 min. at 98°C using the PCR machine and immediately cool it to 4°C.

3. Add probe to prehybridization solution (25 ng/ml; 30 ml = 750 ng total probe) and mix well but avoid foaming. Bubbles may lead to background.

4. Pour off the prehybridization solution from the hybridization bags and add new prehybridization and probe solution mixture to the bags containing the membrane.

5. Incubate with gentle agitation for at least 16 hours.

6. Proceed to medium stringency post-hybridization wash:

Three times for 20 min. each with gentle agitation using 1X SSC, 1% SDS at 60°C.

All wash solutions must be prewarmed to 60°C. Use about 100 ml of wash solution per membrane.

7. To avoid background keep the membranes fully submerged to avoid drying in spots; agitate sufficiently to avoid having membranes stick to one another.

After the wash, proceed to immunological detection and CSPD development.

#### **E. Procedure for Immunological Detection with CSPD**

##### Solutions:

Buffer 1: Maleic acid buffer (0.1 M maleic acid, 0.15 M NaCl; adjusted to pH 7.5 with NaOH)

Washing buffer: Maleic acid buffer with 0.3% (v/v) Tween 20.

09679203 1004100

Blocking stock solution 10% blocking reagent in buffer 1. Dissolve (10X concentration): blocking reagent powder (Boehringer Mannheim, Indianapolis, IN, cat. no. 1096176) by constantly stirring on a 65°C heating block or heat in a microwave, autoclave and store at 4°C.

Buffer 2  
(1X blocking solution): Dilute the stock solution 1:10 in Buffer 1.

Detection buffer: 0.1 M Tris, 0.1 M NaCl, pH 9.5

Procedure:

1. After the post-hybridization wash the blots are briefly rinsed (1-5 min.) in the maleate washing buffer with gentle shaking.
2. Then the membranes are incubated for 30 min. in Buffer 2 with gentle shaking.
3. Anti-DIG-AP conjugate (Boehringer Mannheim, Indianapolis, IN, cat. no. 1093274) at 75 mU/ml (1:10,000) in Buffer 2 is used for detection. 75 ml of solution can be used for 3 blots.
4. The membrane is incubated for 30 min. in the antibody solution with gentle shaking.
5. The membrane are washed twice in washing buffer with gentle shaking. About 250 mls is used per wash for 3 blots.
6. The blots are equilibrated for 2-5 min in 60 ml detection buffer.
7. Dilute CSPD (1:200) in detection buffer. (This can be prepared ahead of time and stored in the dark at 4°C).

The following steps must be done individually. Bags (one for detection and one for exposure) are generally cut and ready before doing the following steps.

00400T 00267960

8. The blot is carefully removed from the detection buffer and excess liquid removed without drying the membrane. The blot is immediately placed in a bag and 1.5 ml of CSPD solution is added. The CSPD solution can be spread over the membrane. Bubbles present at the edge and on the surface of the blot are typically removed by gentle rubbing. The membrane is incubated for 5 min. in CSPD solution.
9. Excess liquid is removed and the membrane is blotted briefly (DNA side up) on Whatman 3MM paper. Do not let the membrane dry completely.
10. Seal the damp membrane in a hybridization bag and incubate for 10 min at 37°C to enhance the luminescent reaction.
11. Expose for 2 hours at room temperature to X-ray film. Multiple exposures can be taken. Luminescence continues for at least 24 hours and signal intensity increases during the first hours.

### **Example 3: Transformation of Carrot Cells**

Transformation of plant cells can be accomplished by a number of methods, as described above. Similarly, a number of plant genera can be regenerated from tissue culture following transformation. Transformation and regeneration of carrot cells as described herein is illustrative.

Single cell suspension cultures of carrot (*Daucus carota*) cells are established from hypocotyls of cultivar Early Nantes in B<sub>5</sub> growth medium (O.L. Gamborg et al., *Plant Physiol.* 45:372 (1970)) plus 2,4-D and 15 mM CaCl<sub>2</sub> (B<sub>5</sub>-44 medium) by methods known in the art. The suspension cultures are subcultured by adding 10 ml of the suspension culture to 40 ml of B<sub>5</sub>-44 medium in 250 ml flasks every 7 days and are maintained in a shaker at 150 rpm at 27 °C in the dark.

The suspension culture cells are transformed with exogenous DNA as described by Z. Chen et al. *Plant Mol. Bio.* 36:163 (1998). Briefly, 4-days post-subculture cells are incubated with cell wall digestion solution containing 0.4 M sorbitol, 2% driselase, 5mM MES (2-[N-Morpholino] ethanesulfonic acid) pH 5.0 for 5 hours. The digested cells are pelleted gently at 60 xg for 5 min. and washed twice in W5 solution containing 154 mM NaCl, 5 mM KCl, 125 mM CaCl<sub>2</sub> and 5mM glucose, pH 6.0. The protoplasts are suspended in MC solution



containing 5 mM MES, 20 mM CaCl<sub>2</sub>, 0.5 M mannitol, pH 5.7 and the protoplast density is adjusted to about  $4 \times 10^6$  protoplasts per ml.

15-60 µg of plasmid DNA is mixed with 0.9 ml of protoplasts. The resulting suspension is mixed with 40% polyethylene glycol (MW 8000, PEG 8000), by gentle inversion a few times at room temperature for 5 to 25 min. Protoplast culture medium known in the art is added into the PEG-DNA-protoplast mixture. Protoplasts are incubated in the culture medium for 24 hour to 5 days and cell extracts can be used for assay of transient expression of the introduced gene. Alternatively, transformed cells can be used to produce transgenic callus, which in turn can be used to produce transgenic plants, by methods known in the art. See, for example, Nomura and Komamine, *Plt. Phys.* 79:988-991 (1985), *Identification and Isolation of Single Cells that Produce Somatic Embryos in Carrot Suspension Cultures*.

An additional deposit, PTA-1411, of an *E. coli* Library, *E. coli*LibA021800, was made at the American Type Culture Collection in Manassas, Virginia, USA on February 22, 2000 to meet the requirements of Budapest Treaty for the international recognition of the deposit of microorganisms. This deposit was assigned ATCC accession no. PTA-1411.

The invention being thus described, it will be apparent to one of ordinary skill in the art that various modifications of the materials and methods for practicing the invention can be made. Such modifications are to be considered within the scope of the invention as defined by the following claims.

Each of the references from the patent and periodical literature cited herein is hereby expressly incorporated in its entirety by such citation.

004007 00264960

Table 1

| >6006847 / |        |        |       |     |   |
|------------|--------|--------|-------|-----|---|
|            | len =  | 101410 | nex = | 200 |   |
| 5          | Intr   | 101    | 172   | +   | 1 |
|            | Intr   | 302    | 396   | +   | 1 |
|            | Intr   | 465    | 531   | +   | 1 |
|            | Intr   | 615    | 744   | +   | 1 |
| 10         | Intr   | 827    | 1045  | +   | 1 |
|            | Intr   | 1129   | 1259  | +   | 1 |
|            | Intr   | 1556   | 1730  | +   | 1 |
|            | Intr   | 1821   | 1911  | +   | 1 |
|            | Intr   | 2009   | 2166  | +   | 1 |
| 15         | Intr   | 2262   | 2450  | +   | 1 |
|            | Intr   | 2674   | 2733  | +   | 1 |
|            | Intr   | 3297   | 3374  | +   | 1 |
|            | Intr   | 3687   | 3792  | +   | 1 |
|            | Term   | 4121   | 4131  | +   | 1 |
| 20         | Init   | 4221   | 4285  | +   | 2 |
|            | Intr   | 4374   | 4502  | +   | 2 |
|            | Intr   | 4588   | 4762  | +   | 2 |
|            | Intr   | 4853   | 4943  | +   | 2 |
| 25         | Intr   | 5021   | 5178  | +   | 2 |
|            | Intr   | 5257   | 5445  | +   | 2 |
|            | Term   | 5537   | 5593  | +   | 2 |
|            | Init   | 6652   | 6829  | +   | 3 |
| 30         | Intr   | 7207   | 7355  | +   | 3 |
|            | Intr   | 7433   | 7548  | +   | 3 |
|            | Term   | 7743   | 7908  | +   | 3 |
|            | Term   | 8370   | 8185  | -   | 4 |
| 35         | Intr   | 8544   | 8458  | -   | 4 |
|            | Intr   | 8706   | 8631  | -   | 4 |
|            | Intr   | 9041   | 8791  | -   | 4 |
|            | Intr   | 9232   | 9116  | -   | 4 |
|            | Intr   | 9410   | 9336  | -   | 4 |
| 40         | Intr   | 9714   | 9493  | -   | 4 |
|            | Intr   | 9912   | 9799  | -   | 4 |
|            | Init   | 10118  | 10011 | -   | 4 |
|            | Single | 12038  | 12232 | +   | 5 |
| 45         | Init   | 12760  | 12883 | +   | 6 |
|            | Intr   | 12994  | 13143 | +   | 6 |
|            | Intr   | 13183  | 13269 | +   | 6 |
|            | Intr   | 13397  | 13542 | +   | 6 |
| 50         | Intr   | 13609  | 13693 | +   | 6 |
|            | Intr   | 13804  | 13850 | +   | 6 |
|            | Intr   | 13955  | 14036 | +   | 6 |
|            | Term   | 14493  | 14779 | +   | 6 |
| 55         | Init   | 15035  | 15196 | +   | 7 |
|            | Intr   | 15533  | 15595 | +   | 7 |
|            | Intr   | 15703  | 15789 | +   | 7 |
|            | Intr   | 15943  | 16041 | +   | 7 |
|            | Intr   | 16129  | 16224 | +   | 7 |
| 60         | Intr   | 16303  | 16368 | +   | 7 |

004007 0026960

1075

004007" 00264960

|    |        |       |       |   |    |
|----|--------|-------|-------|---|----|
|    | Intr   | 16477 | 16707 | + | 7  |
|    | Intr   | 16788 | 16853 | + | 7  |
|    | Intr   | 16991 | 17041 | + | 7  |
|    | Intr   | 17126 | 17190 | + | 7  |
| 5  | Term   | 17313 | 17364 | + | 7  |
|    | Term   | 18055 | 18002 | - | 8  |
|    | Intr   | 18253 | 18176 | - | 8  |
|    | Intr   | 18652 | 18360 | - | 8  |
| 10 | Init   | 18908 | 18752 | - | 8  |
|    | Term   | 19967 | 19961 | - | 9  |
|    | Init   | 21511 | 20829 | - | 9  |
| 15 | Term   | 23137 | 22716 | - | 10 |
|    | Intr   | 23400 | 23276 | - | 10 |
|    | Intr   | 24007 | 23826 | - | 10 |
|    | Init   | 24566 | 24393 | - | 10 |
| 20 | Term   | 27663 | 27657 | - | 11 |
|    | Intr   | 28142 | 27844 | - | 11 |
|    | Intr   | 28355 | 28243 | - | 11 |
|    | Intr   | 29372 | 28495 | - | 11 |
|    | Init   | 29979 | 29504 | - | 11 |
| 25 | Init   | 32376 | 32405 | + | 12 |
|    | Term   | 33427 | 33504 | + | 12 |
| 30 | Single | 33834 | 33553 | - | 13 |
|    | Single | 34444 | 34839 | + | 14 |
|    | Single | 35773 | 35486 | - | 15 |
| 35 | Single | 36634 | 36912 | + | 16 |
|    | Init   | 39808 | 40074 | + | 17 |
|    | Intr   | 40685 | 40767 | + | 17 |
|    | Intr   | 41096 | 41294 | + | 17 |
| 40 | Intr   | 41376 | 41460 | + | 17 |
|    | Intr   | 41781 | 41851 | + | 17 |
|    | Intr   | 41937 | 42025 | + | 17 |
|    | Intr   | 42154 | 42397 | + | 17 |
|    | Intr   | 42608 | 42720 | + | 17 |
| 45 | Intr   | 43007 | 43097 | + | 17 |
|    | Intr   | 43243 | 43371 | + | 17 |
|    | Intr   | 43644 | 44197 | + | 17 |
|    | Intr   | 44412 | 44793 | + | 17 |
|    | Intr   | 44892 | 44928 | + | 17 |
| 50 | Term   | 45011 | 45162 | + | 17 |
|    | Term   | 45744 | 45571 | - | 18 |
|    | Intr   | 45893 | 45828 | - | 18 |
|    | Intr   | 46050 | 45970 | - | 18 |
| 55 | Intr   | 46198 | 46125 | - | 18 |
|    | Intr   | 46465 | 46405 | - | 18 |
|    | Intr   | 46611 | 46541 | - | 18 |
|    | Intr   | 46738 | 46696 | - | 18 |
|    | Intr   | 46915 | 46827 | - | 18 |
| 60 | Intr   | 47121 | 47016 | - | 18 |
|    | Intr   | 47313 | 47208 | - | 18 |

1076

004001 60262960

|    |        |       |       |   |    |
|----|--------|-------|-------|---|----|
|    | Intr   | 47748 | 47701 | - | 18 |
|    | Intr   | 47911 | 47842 | - | 18 |
|    | Intr   | 48181 | 48145 | - | 18 |
|    | Intr   | 49398 | 49195 | - | 18 |
| 5  | Intr   | 50057 | 49515 | - | 18 |
|    | Intr   | 50822 | 50166 | - | 18 |
|    | Intr   | 51452 | 50955 | - | 18 |
|    | Intr   | 51762 | 51550 | - | 18 |
|    | Intr   | 52202 | 51913 | - | 18 |
| 10 | Intr   | 52533 | 52394 | - | 18 |
|    | Intr   | 53068 | 52996 | - | 18 |
|    | Intr   | 53355 | 53178 | - | 18 |
|    | Init   | 53836 | 53564 | - | 18 |
| 15 | Term   | 56659 | 56291 | - | 19 |
|    | Intr   | 57188 | 56736 | - | 19 |
|    | Intr   | 57376 | 57271 | - | 19 |
|    | Intr   | 57636 | 57470 | - | 19 |
|    | Intr   | 57824 | 57744 | - | 19 |
| 20 | Intr   | 58074 | 57911 | - | 19 |
|    | Intr   | 58571 | 58180 | - | 19 |
|    | Intr   | 58795 | 58667 | - | 19 |
|    | Intr   | 58979 | 58893 | - | 19 |
|    | Init   | 59429 | 59095 | - | 19 |
| 25 | Init   | 61808 | 61831 | + | 20 |
|    | Intr   | 61923 | 61998 | + | 20 |
|    | Term   | 62406 | 62887 | + | 20 |
| 30 | Term   | 63682 | 63453 | - | 21 |
|    | Intr   | 63845 | 63800 | - | 21 |
|    | Intr   | 63968 | 63940 | - | 21 |
|    | Intr   | 64321 | 64165 | - | 21 |
|    | Intr   | 64643 | 64463 | - | 21 |
| 35 | Init   | 65146 | 65142 | - | 21 |
|    | Init   | 68595 | 69930 | + | 22 |
|    | Intr   | 70025 | 70158 | + | 22 |
|    | Intr   | 70252 | 70492 | + | 22 |
| 40 | Intr   | 70589 | 70794 | + | 22 |
|    | Intr   | 70868 | 70911 | + | 22 |
|    | Intr   | 71080 | 71191 | + | 22 |
|    | Intr   | 71271 | 71583 | + | 22 |
|    | Intr   | 71706 | 71724 | + | 22 |
| 45 | Term   | 71836 | 71860 | + | 22 |
|    | Term   | 72768 | 72358 | - | 23 |
|    | Init   | 73399 | 72989 | - | 23 |
| 50 | Term   | 75920 | 75520 | - | 24 |
|    | Init   | 77055 | 76392 | - | 24 |
|    | Init   | 77311 | 77678 | + | 25 |
|    | Intr   | 78802 | 79147 | + | 25 |
| 55 | Intr   | 79247 | 79529 | + | 25 |
|    | Term   | 79612 | 79925 | + | 25 |
|    | Single | 80711 | 81892 | + | 26 |
| 60 | Init   | 82567 | 82807 | + | 27 |
|    | Intr   | 82918 | 83174 | + | 27 |

1077

|    |          |        |        |     |    |
|----|----------|--------|--------|-----|----|
|    | Intr     | 83285  | 83592  | +   | 27 |
|    | Intr     | 83719  | 83800  | +   | 27 |
|    | Term     | 84172  | 84249  | +   | 27 |
| 5  | Init     | 84338  | 84373  | +   | 28 |
|    | Intr     | 84462  | 84587  | +   | 28 |
|    | Term     | 84668  | 84745  | +   | 28 |
| 10 | Init     | 85543  | 85765  | +   | 29 |
|    | Intr     | 86073  | 86726  | +   | 29 |
|    | Intr     | 87131  | 87439  | +   | 29 |
|    | Intr     | 87526  | 87727  | +   | 29 |
|    | Intr     | 87808  | 87892  | +   | 29 |
|    | Intr     | 88123  | 88263  | +   | 29 |
| 15 | Intr     | 88352  | 88470  | +   | 29 |
|    | Intr     | 88557  | 88653  | +   | 29 |
|    | Intr     | 88853  | 89025  | +   | 29 |
|    | Intr     | 89144  | 89210  | +   | 29 |
|    | Intr     | 89362  | 89520  | +   | 29 |
| 20 | Intr     | 89743  | 89947  | +   | 29 |
|    | Intr     | 90079  | 90187  | +   | 29 |
|    | Intr     | 90481  | 90593  | +   | 29 |
|    | Intr     | 90858  | 90933  | +   | 29 |
|    | Term     | 91014  | 91161  | +   | 29 |
| 25 | Term     | 92436  | 92125  | -   | 30 |
|    | Intr     | 92576  | 92535  | -   | 30 |
|    | Intr     | 92783  | 92666  | -   | 30 |
|    | Intr     | 92954  | 92860  | -   | 30 |
| 30 | Intr     | 93145  | 93038  | -   | 30 |
|    | Intr     | 93307  | 93234  | -   | 30 |
|    | Intr     | 93504  | 93432  | -   | 30 |
|    | Intr     | 93659  | 93612  | -   | 30 |
|    | Intr     | 93919  | 93736  | -   | 30 |
| 35 | Intr     | 94194  | 94016  | -   | 30 |
|    | Intr     | 94448  | 94356  | -   | 30 |
|    | Intr     | 94692  | 94579  | -   | 30 |
|    | Intr     | 94995  | 94882  | -   | 30 |
|    | Intr     | 95364  | 95101  | -   | 30 |
| 40 | Intr     | 95579  | 95465  | -   | 30 |
|    | Intr     | 95719  | 95664  | -   | 30 |
|    | Intr     | 95982  | 95826  | -   | 30 |
|    | Intr     | 96139  | 96075  | -   | 30 |
|    | Intr     | 96489  | 96406  | -   | 30 |
| 45 | Intr     | 96610  | 96568  | -   | 30 |
|    | Intr     | 96817  | 96711  | -   | 30 |
|    | Intr     | 97013  | 96945  | -   | 30 |
|    | Init     | 97489  | 97196  | -   | 30 |
| 50 | Term     | 98901  | 98840  | -   | 31 |
|    | Intr     | 100135 | 99961  | -   | 31 |
|    | Intr     | 100583 | 100427 | -   | 31 |
|    | Init     | 100733 | 100729 | -   | 31 |
| 55 | >6006873 | /      |        |     |    |
|    | len =    | 117296 | nex =  | 155 |    |
|    | Term     | 42     | 65     | +   | 1  |
| 60 | Init     | 550    | 631    | +   | 2  |
|    | Intr     | 1004   | 1191   | +   | 2  |
|    | Intr     | 1279   | 1438   | +   | 2  |

0969203-100400

1078

|    |        |       |       |   |    |
|----|--------|-------|-------|---|----|
| 5  | Intr   | 1506  | 1808  | + | 2  |
|    | Intr   | 1848  | 1971  | + | 2  |
|    | Intr   | 2003  | 2164  | + | 2  |
|    | Intr   | 2304  | 2573  | + | 2  |
|    | Term   | 2675  | 2858  | + | 2  |
|    | Init   | 3646  | 5335  | + | 3  |
|    | Term   | 5767  | 5972  | + | 3  |
| 10 | Init   | 6614  | 7403  | + | 4  |
|    | Term   | 7497  | 8314  | + | 4  |
|    | Single | 9113  | 11413 | + | 5  |
| 15 | Single | 12363 | 14495 | + | 6  |
| 20 | Init   | 16051 | 18031 | + | 7  |
|    | Intr   | 18679 | 18865 | + | 7  |
|    | Intr   | 18949 | 18989 | + | 7  |
|    | Intr   | 19122 | 19200 | + | 7  |
|    | Intr   | 19693 | 19779 | + | 7  |
|    | Intr   | 19855 | 19943 | + | 7  |
|    | Intr   | 20038 | 20109 | + | 7  |
| 25 | Term   | 20197 | 20300 | + | 7  |
|    | Init   | 21071 | 21443 | + | 8  |
|    | Term   | 22354 | 22901 | + | 8  |
| 30 | Single | 23430 | 23894 | + | 9  |
|    | Term   | 27029 | 26716 | - | 10 |
|    | Intr   | 27227 | 27122 | - | 10 |
|    | Intr   | 27655 | 27320 | - | 10 |
| 35 | Init   | 28184 | 27741 | - | 10 |
|    | Init   | 34369 | 35235 | + | 11 |
|    | Intr   | 36115 | 36657 | + | 11 |
|    | Term   | 36751 | 36858 | + | 11 |
| 40 | Init   | 38446 | 39033 | + | 12 |
|    | Term   | 39293 | 40213 | + | 12 |
| 45 | Init   | 42013 | 42031 | + | 13 |
|    | Intr   | 43446 | 43521 | + | 13 |
|    | Term   | 44625 | 44769 | + | 13 |
| 50 | Init   | 45591 | 45962 | + | 14 |
|    | Intr   | 46082 | 47090 | + | 14 |
|    | Term   | 47160 | 47464 | + | 14 |
|    | Single | 49937 | 47994 | - | 15 |
|    | Single | 50785 | 52656 | + | 16 |
| 55 | Term   | 52875 | 52796 | - | 17 |
|    | Intr   | 53092 | 52963 | - | 17 |
|    | Intr   | 53344 | 53240 | - | 17 |
|    | Intr   | 53585 | 53430 | - | 17 |
| 60 | Intr   | 53785 | 53684 | - | 17 |
|    | Intr   | 53937 | 53875 | - | 17 |
|    | Intr   | 54439 | 54259 | - | 17 |

09679203 100400

00400T" E0262960

|    |      |       |       |   |    |
|----|------|-------|-------|---|----|
|    | Init | 54614 | 54544 | - | 17 |
|    | Init | 57079 | 57116 | + | 18 |
|    | Intr | 57231 | 57393 | + | 18 |
| 5  | Intr | 58416 | 58814 | + | 18 |
|    | Intr | 58908 | 59165 | + | 18 |
|    | Term | 59331 | 59672 | + | 18 |
|    | Term | 60498 | 60058 | - | 19 |
| 10 | Intr | 60805 | 60575 | - | 19 |
|    | Intr | 61020 | 60901 | - | 19 |
|    | Intr | 61237 | 61117 | - | 19 |
|    | Init | 62102 | 61708 | - | 19 |
| 15 | Term | 64220 | 64161 | - | 20 |
|    | Intr | 64430 | 64309 | - | 20 |
|    | Intr | 64839 | 64817 | - | 20 |
|    | Init | 65009 | 64975 | - | 20 |
| 20 | Term | 66359 | 66351 | - | 21 |
|    | Intr | 66585 | 66456 | - | 21 |
|    | Intr | 66753 | 66668 | - | 21 |
|    | Intr | 66896 | 66851 | - | 21 |
|    | Intr | 67102 | 66984 | - | 21 |
| 25 | Init | 67401 | 67399 | - | 21 |
|    | Term | 68154 | 68135 | - | 22 |
|    | Intr | 68830 | 68244 | - | 22 |
|    | Intr | 69023 | 68922 | - | 22 |
| 30 | Intr | 69474 | 69104 | - | 22 |
|    | Init | 69674 | 69636 | - | 22 |
|    | Term | 71519 | 71336 | - | 23 |
|    | Intr | 71723 | 71599 | - | 23 |
| 35 | Intr | 72651 | 72545 | - | 23 |
|    | Init | 72859 | 72841 | - | 23 |
|    | Init | 75377 | 75772 | + | 24 |
|    | Intr | 75930 | 76091 | + | 24 |
| 40 | Intr | 76178 | 76353 | + | 24 |
|    | Intr | 76459 | 76534 | + | 24 |
|    | Intr | 76741 | 76782 | + | 24 |
|    | Intr | 76823 | 76951 | + | 24 |
|    | Intr | 77212 | 77331 | + | 24 |
| 45 | Intr | 77414 | 77482 | + | 24 |
|    | Intr | 77567 | 77648 | + | 24 |
|    | Intr | 77822 | 77916 | + | 24 |
|    | Intr | 77997 | 78061 | + | 24 |
|    | Intr | 78142 | 78340 | + | 24 |
| 50 | Intr | 78458 | 78652 | + | 24 |
|    | Intr | 78747 | 78841 | + | 24 |
|    | Intr | 78937 | 79072 | + | 24 |
|    | Intr | 79167 | 79384 | + | 24 |
|    | Intr | 79460 | 79587 | + | 24 |
| 55 | Term | 79679 | 80082 | + | 24 |
|    | Init | 80797 | 81071 | + | 25 |
|    | Intr | 81162 | 81285 | + | 25 |
|    | Term | 81372 | 81587 | + | 25 |
| 60 | Init | 82088 | 82128 | + | 26 |

1080

|    |        |        |        |   |    |
|----|--------|--------|--------|---|----|
|    | Intr   | 82408  | 82843  | + | 26 |
|    | Intr   | 82937  | 83028  | + | 26 |
|    | Intr   | 83459  | 83623  | + | 26 |
|    | Intr   | 83753  | 84146  | + | 26 |
| 5  | Intr   | 84500  | 84618  | + | 26 |
|    | Intr   | 84722  | 84794  | + | 26 |
|    | Intr   | 84872  | 85098  | + | 26 |
|    | Intr   | 85190  | 85234  | + | 26 |
|    | Intr   | 85583  | 85625  | + | 26 |
| 10 | Intr   | 85728  | 85952  | + | 26 |
|    | Intr   | 86166  | 86269  | + | 26 |
|    | Intr   | 86394  | 86460  | + | 26 |
|    | Intr   | 86681  | 86830  | + | 26 |
|    | Intr   | 87178  | 87300  | + | 26 |
| 15 | Intr   | 87407  | 87577  | + | 26 |
|    | Intr   | 87782  | 87841  | + | 26 |
|    | Term   | 87925  | 88041  | + | 26 |
|    | Term   | 88225  | 88092  | - | 27 |
| 20 | Intr   | 88402  | 88269  | - | 27 |
|    | Intr   | 88650  | 88492  | - | 27 |
|    | Intr   | 89138  | 88749  | - | 27 |
|    | Init   | 89863  | 89457  | - | 27 |
|    | Term   | 91753  | 91700  | - | 28 |
| 25 | Intr   | 92249  | 92193  | - | 28 |
|    | Intr   | 92587  | 92415  | - | 28 |
|    | Init   | 93806  | 93230  | - | 28 |
|    | Init   | 97637  | 97675  | + | 29 |
| 30 | Intr   | 98393  | 98531  | + | 29 |
|    | Term   | 103138 | 103160 | + | 29 |
|    | Init   | 103163 | 103645 | + | 30 |
| 35 | Intr   | 104321 | 104443 | + | 30 |
|    | Intr   | 104810 | 105013 | + | 30 |
|    | Intr   | 105833 | 106068 | + | 30 |
|    | Term   | 106853 | 106997 | + | 30 |
| 40 | Single | 109872 | 110225 | + | 31 |
|    | Term   | 112229 | 112086 | - | 32 |
|    | Intr   | 112424 | 112320 | - | 32 |
|    | Intr   | 112645 | 112595 | - | 32 |
| 45 | Intr   | 112952 | 112805 | - | 32 |
|    | Intr   | 113126 | 113044 | - | 32 |
|    | Intr   | 113283 | 113224 | - | 32 |
|    | Intr   | 113463 | 113413 | - | 32 |
|    | Intr   | 113620 | 113570 | - | 32 |
| 50 | Intr   | 113954 | 113731 | - | 32 |
|    | Init   | 114105 | 114036 | - | 32 |
|    | Init   | 114724 | 114930 | + | 33 |
|    | Intr   | 115314 | 115388 | + | 33 |
| 55 | Intr   | 115550 | 115631 | + | 33 |
|    | Intr   | 115727 | 115806 | + | 33 |
|    | Intr   | 115931 | 116056 | + | 33 |
|    | Intr   | 116151 | 116264 | + | 33 |
|    | Term   | 116356 | 116472 | + | 33 |
| 60 | Term   | 117296 | 117144 | - | 34 |

004001 00262960



# SCHEMATIC OF A GENE

